

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 09:56:54 ; Search time 38.59 seconds
(without alignments)
3373.374 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWSPSTQAGH.....FVFSQEMVYFSLKVECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------|----------------------|
| 1 | 6605 | 100.0 | 1172 | 21 AAB19677 | Human thrombospondin |
| 2 | 6570 | 99.5 | 1172 | 21 AAB00043 | Human thrombospondin |
| 3 | 4228 | 64.0 | 1170 | 22 AAB90800 | Human shear stress |
| 4 | 4228 | 64.0 | 1170 | 22 AAB74450 | Human variant thro |
| 5 | 4226 | 64.0 | 1152 | 21 AAB00042 | Human thrombospondin |
| 6 | 3251 | 49.2 | 571 | 22 AAG75130 | Human colon cancer |
| 7 | 3236 | 49.0 | 571 | 21 AAB43386 | Human cancer assoc |
| 8 | 2094 | 31.7 | 731 | 22 AAU02913 | Angiotensin conver |
| 9 | 1990 | 30.1 | 757 | 21 AAB00044 | Human cartilage ol |
| 10 | 1946.5 | 29.5 | 762 | 22 AAM93335 | Human polypeptide, |
| 11 | 1899 | 28.8 | 899 | 15 AAR56248 | Xenopus thrombospo |

| | | | | | |
|----|--------|------|------|-------------|----------------------|
| 12 | 1890 | 28.6 | 961 | 15 AAR56249 | Human thrombospondin |
| 13 | 1805 | 27.3 | 1024 | 22 AAB56849 | Drosophila melanog |
| 14 | 1428 | 21.6 | 444 | 21 AAB58933 | Breast and ovarian |
| 15 | 1428 | 21.6 | 444 | 22 AAG73734 | Human colon cancer |
| 16 | 1359 | 20.6 | 555 | 22 AAU02914 | Angiotensin conver |
| 17 | 1302 | 19.7 | 776 | 22 AAM79078 | Human protein SEQ |
| 18 | 1193 | 18.1 | 541 | 22 AAB74451 | Human variant thro |
| 19 | 1123 | 17.0 | 546 | 22 AAU02915 | Angiotensin conver |
| 20 | 1012.5 | 15.3 | 300 | 21 AAB00041 | Human COMP/TSP-2 c |
| 21 | 880 | 13.3 | 239 | 14 AAR40823 | Human thrombospondin |
| 22 | 850 | 12.9 | 459 | 22 AAU02916 | Angiotensin conver |
| 23 | 817 | 12.4 | 441 | 19 AAM40288 | Human concatameris |
| 24 | 783 | 11.9 | 441 | 20 AAY06183 | Thrombospondin I f |
| 25 | 779 | 11.8 | 466 | 21 AAB43602 | Human cancer assoc |
| 26 | 750.5 | 11.4 | 622 | 22 AAM80062 | Human protein SEQ |
| 27 | 748 | 11.3 | 218 | 19 AAM40287 | Human TSP1 protein |
| 28 | 713 | 10.8 | 218 | 20 AAY06182 | Thrombospondin I f |
| 29 | 645 | 9.8 | 157 | 21 AAB08133 | Amino acid sequenc |
| 30 | 612 | 9.3 | 151 | 21 AAB53711 | Human colon cancer |
| 31 | 513.5 | 7.8 | 242 | 21 AAB00040 | Human COMP/TSP-1 c |
| 32 | 507 | 7.7 | 1688 | 22 AAG67244 | Amino acid sequenc |
| 33 | 476.5 | 7.2 | 877 | 22 AAU16959 | Human novel secret |
| 34 | 472.5 | 7.2 | 800 | 22 ABB10492 | Human cDNA SEQ ID |
| 35 | 472.5 | 7.2 | 800 | 22 AAB18148 | Novel human uterin |
| 36 | 472.5 | 7.2 | 800 | 22 AAU17031 | Human novel secret |
| 37 | 424 | 6.4 | 1584 | 20 AAM99300 | Human BAI3 protein |
| 38 | 418 | 6.3 | 1522 | 20 AAM99302 | Human BAI3 protein |
| 39 | 418 | 6.3 | 1522 | 21 AAB23601 | Human secreted pro |
| 40 | 415 | 6.3 | 1075 | 22 AAU32291 | Novel human secret |
| 41 | 400 | 6.1 | 999 | 21 AAY94990 | Human secreted pro |
| 42 | 390 | 5.9 | 1093 | 22 AAF02455 | Mouse semaphorin G |
| 43 | 388.5 | 5.9 | 810 | 18 AAM37500 | Human nel-related |
| 44 | 384.5 | 5.8 | 1572 | 20 AAM99301 | Human BAI2 protein |
| 45 | 362 | 5.8 | 780 | 22 AAG67241 | Amino acid sequenc |

ALIGNMENTS

RESULT 1

AAB19677
ID AAB19677 standard; Protein; 1172 AA.

XX AAB19677;

XX 05-FEB-2001 (first entry)

XX Human thrombospondin-2.

XX Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
XX melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
XX prostate cancer; psoriasis; rosecea dermatosis; antitumour;
XX therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 382..429

FT Region /note= "type 1 repeat"

FT Region 384..390

FT Region /note= "WSPNAEW sequence involved in antiangiogenic activity"

FT Region 438..490

FT Region /note= "type 1 repeat"

FT Region 495..547

FT Region /note= "type 1 repeat"

PN WO200057899-A1.

XX 05-OCT-2000.

XX 24-MAR-2000; 2000WO-US07835.

XX

PR 31-MAR-1999; 99US-0127221.
XX (GEO) GEN HOSPITAL CORP.
XX
PI Detmar M, Streit M;
XX
DR WPI; 2000-656131/63.
DR N-PSDB; AAA88669.
XX
PT Treating a disorder characterized by unwanted cell proliferation e.g. precancerous, cancerous or neoplastic cells or presence of tumor
PT preferably of skin or prostate, comprises increasing thrombospondin-2
PT activity
XX
PS Disclosure; Fig 2; 73pp; English.
XX
CC The present sequence is that of human thrombospondin-2 (TSP-2).
CC The invention is based on the discovery that overexpression of
CC TSP-2 decreases tumor size in vivo, and features methods for
CC modulating unwanted angiogenesis and tumour growth. Treatment of
CC unwanted cell proliferation or angiogenesis involves increasing
CC TSP-2 activity. This is achieved by administering an agent which
CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
CC derived polypeptide or retro-inverso peptide, a nucleic acid
CC encoding TSP-2, an agonist of TSP-2, or an agent that increases
CC TSP-2 gene expression. The TSP-2 polypeptide may include at least
CC one type 1 repeat such as the WSPWAEW peptide (see AAB19683).
CC which is involved in the antiangiogenic activity of TSP-2. The
CC method is used to treat a disorder characterised by pre-cancerous,
CC cancerous or neoplastic cells, or the presence of a tumour, or a
CC disorder that affects epithelial tissues resulting in unwanted
CC skin cell proliferation. Such disorders include malignant
CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,
CC rosacea dermatosis, psoriasis, and skin damage caused by
CC photoradiation (all claimed). Evaluating the presence of TSP-2
CC nucleic acid or protein is useful for diagnosing a subject at risk
CC of unwanted cell proliferation or angiogenesis. Methods are also
CC provided of identifying compounds that modulate TSP-2 activity.
XX
SQ Sequence 1172 AA;

Query Match 100.0%; Score 6605; DB 21; Length 1172;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYRLVLLALWVPSQAGHQBKDTFDFLSISNIRKTIAGAKQRPDPGPAYRVRVF 60
DB 1 MWRLVLLALWVPSQAGHQBKDTFDFLSISNIRKTIAGAKQRPDPGPAYRVRVF 60

QY 61 DYIPVNNADLSKITKIMRQKSGFELTAQLKQDGKSRGTLALLEGFGLSQRQFEIVSNGP 120
DB 61 DYIPVNNADLSKITKIMRQKSGFELTAQLKQDGKSRGTLALLEGFGLSQRQFEIVSNGP 120

QY 121 ADTLTLTWIDGTRHVSLEVDGLADSQWKNVTVOVAGETSLHVGCDLIDSFALDEPFF 180
DB 121 ADTLTLTWIDGTRHVSLEVDGLADSQWKNVTVOVAGETSLHVGCDLIDSFALDEPFF 180

QY 181 EHLQAEKSMYVAKGASRSHRFGILLQNVHLVFENSVEDILSKKCGQOGAGAEINAISEN 240
DB 181 EHLQAEKSMYVAKGASRSHRFGILLQNVHLVFENSVEDILSKKCGQOGAGAEINAISEN 240

QY 241 TETRLGLPHVTTEYVGFSPSERPEYCERSCSELGNMVELSLHVLVNPQSENLRKRVND 300
DB 241 TETRLGLPHVTTEYVGFSPSERPEYCERSCSELGNMVELSLHVLVNPQSENLRKRVND 300

QY 301 NOFLHELIGGPKTRNMGACWDGRRFAENETWVVDSTCTCTCKFKTICHOITCPPATC 360
DB 301 NOFLHELIGGPKTRNMGACWDGRRFAENETWVVDSTCTCTCKFKTICHOITCPPATC 360

QY 361 ASPSFVEGECPCSLHSDVDEGEGFWAPWTCVTCGSGTQGRSCDVTNTCLGPSI 420
DB 361 ASPSFVEGECPCSLHSDVDEGEGFWAPWTCVTCGSGTQGRSCDVTNTCLGPSI 420

QY 421 QTRACSLSKCDTRIRQDGSWSHSPWSSCVTCGVGNITRILCNPSVPQMGKCKGSG 480
DB 421 QTRACSLSKCDTRIRQDGSWSHSPWSSCVTCGVGNITRILCNPSVPQMGKCKGSG 480

QY 481 RETKACQCAPCIDGRWSPWSPWSSACTVTCAGIRIRTRVCNSPEPQYGGKACVGDVQER 540
DB 481 RETKACQCAPCIDGRWSPWSPWSSACTVTCAGIRIRTRVCNSPEPQYGGKACVGDVQER 540

QY 541 QMCNKRSCFVDCGLSNPCFPQAQSCSPDGSWSCSPVGLNGTHCEDLDECALVPDI 600
DB 541 QMCNKRSCFVDCGLSNPCFPQAQSCSPDGSWSCSPVGLNGTHCEDLDECALVPDI 600

QY 601 CFSTSKVPCVNTQPGFHCPLPCPPRYRGVGNQPVGVLEAAKTEKQVCEPENCKDKTHNCH 660
DB 601 CFSTSKVPCVNTQPGFHCPLPCPPRYRGVGNQPVGVLEAAKTEKQVCEPENCKDKTHNCH 660

QY 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSDLGWPNLNLVCATNATHYCIKDC 720
DB 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSDLGWPNLNLVCATNATHYCIKDC 720

QY 721 PHLNPSGODEPDKGIGDACDDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780
DB 721 PHLNPSGODEPDKGIGDACDDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780

QY 781 YVHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHCDNCPL 840
DB 781 YVHNPAQIDTNNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHCDNCPL 840

QY 841 VHNPDQTDVNDLVGDCQNNEDIDDDGHQNNQDNCPIYISNANQADHQRDGGDADCDPDD 900
DB 841 VHNPDQTDVNDLVGDCQNNEDIDDDGHQNNQDNCPIYISNANQADHQRDGGDADCDPDD 900

QY 901 DNDGVDPDRDRCRLVFNPDQEDLDGDRGDIKDDDFDNDNIPDIDVCPENNAISETDPR 960
DB 901 DNDGVDPDRDRCRLVFNPDQEDLDGDRGDIKDDDFDNDNIPDIDVCPENNAISETDPR 960

QY 961 NFQMVPLDPKGTQIDPNWIRHOGKELVQTANSDPGIAVGDFEGSDVFGTFYVNTDR 1020
DB 961 NFQMVPLDPKGTQIDPNWIRHOGKELVQTANSDPGIAVGDFEGSDVFGTFYVNTDR 1020

QY 1021 DDDYAGFVFGVQSSRRFVVMKQVTOYWEQDPPRAYGSGVSLKVVNSTTGTGHELRN 1080
DB 1021 DDDYAGFVFGVQSSRRFVVMKQVTOYWEQDPPRAYGSGVSLKVVNSTTGTGHELRN 1080

QY 1081 ALWHTGNTPGVORTLWHPDRNIGWKDYTAIRWHLTHRPKTYIRVLVHEGKQVMAADSGPI 1140
DB 1081 ALWHTGNTPGVORTLWHPDRNIGWKDYTAIRWHLTHRPKTYIRVLVHEGKQVMAADSGPI 1140

QY 1141 YDQTYAGGRGLGLFVSQEMVYFSDLYKBCRDI 1172
DB 1141 YDQTYAGGRGLGLFVSQEMVYFSDLYKBCRDI 1172

RESULT 2
AAB00043
ID AAB00043 standard; Protein; 1172 AA.
XX
AC AAB00043;
XX
XX 08-NOV-2000 (first entry)
XX
XX Human thrombospondin-2 (TSP-2).
XX
XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
XX glaucoma.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 381...436
FT Region

| | | | | | | | |
|---|---|---|-----|----|-----|---|-----|
| FT | Region | /label= Type 1 repeat region | 301 | Db | 301 | nqflwlllgppktrmsacwqdgdrffaenetrwvvdctctckkfkfichqitcpcpatc | 360 |
| FT | Region | 437..493 | | | | | |
| FT | Region | /label= Type 1 repeat region | 361 | QY | 361 | ASPSFVEGECPCSLHSVDGEGSWSPAEMWTQCSVTGSGTQQRGRSDVTSNTCLGPSI | 420 |
| FT | Region | 494..550 | | | | | |
| FT | Region | /label= Type 1 repeat region | 361 | Db | 361 | aspsfvegeccpsclhsvdgeegswspaewtqcsvtcgsgtqqrgrsdvtsntclgpsl | 420 |
| XX | WO200044908-A2. | | | | | | |
| PN | | | | | | | |
| XX | | | | | | | |
| PD | 03-AUG-2000. | | | | | | |
| XX | | | | | | | |
| PF | 01-FEB-2000; 2000WO-US02482. | | | | | | |
| XX | | | | | | | |
| PR | 01-FEB-1999; 99US-0118053. | | | | | | |
| XX | | | | | | | |
| PA | (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT. | | | | | | |
| XX | | | | | | | |
| PI | Lawler JW; | | | | | | |
| XX | | | | | | | |
| DR | WPI; 2000-514823/46. | | | | | | |
| XX | | | | | | | |
| PT | Nucleic acids encoding chimeric proteins such as cartilage oligomeric | | | | | | |
| PT | matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for | | | | | | |
| PT | Inhibiting angiogenesis and treating diseases such as cancer | | | | | | |
| XX | | | | | | | |
| PS | Disclosure; Fig 2; 40pp; English. | | | | | | |
| XX | | | | | | | |
| CC | New nucleic acids are described which encode a protein comprising | | | | | | |
| CC | the second and third type 1 repeats of human TSP (thrombospondin)-1, | | | | | | |
| CC | but not the TGF (transforming growth factor)-beta activation region | | | | | | |
| CC | of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing | | | | | | |
| CC | the second and third type 1 repeats and the COMP (cartilage | | | | | | |
| CC | oligomeric matrix protein) assembly sequence (COMP/TSP-1) was | | | | | | |
| CC | produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1 | | | | | | |
| CC | caused inhibition of the growth of tumours in mice models. | | | | | | |
| CC | Thus the nucleic acids and proteins may be useful for treating | | | | | | |
| CC | angiogenesis related diseases such as cancer (by reducing the rate of | | | | | | |
| CC | growth and size of tumours), arthritis, psoriasis, diabetic | | | | | | |
| CC | retinopathy, corneal graft rejection, and glaucoma. They may also be | | | | | | |
| CC | used for treating human immunodeficiency virus (HIV) infection. | | | | | | |
| CC | Anti-angiogenic therapy has little toxicity, does not require the | | | | | | |
| CC | therapeutic agent to enter tumour cells or cross the blood-brain | | | | | | |
| CC | barrier, controls tumour growth independently of growth of | | | | | | |
| CC | tumour cell heterogeneity, and does not induce drug resistance. | | | | | | |
| XX | | | | | | | |
| SQ | Sequence 1172 AA; | | | | | | |
| <p>Query Match 99.5%; Score 6570; DB 21; Length 1172; Best Local Similarity 99.6%; Pred. No. 0; Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;</p> | | | | | | | |
| QY | 1 | MVRLVLLALWWPSTQAGHODKDTTDLFSISNINRKTIGAKOFGGPDGPVAYRVR | 60 | | | | |
| Db | 1 | mvrlvllalwwpstadqaghdkttdlfsisnlnrktigakqgrpdgpvayrvrf | 60 | | | | |
| QY | 61 | DYTPPVNADDLKITKLMRQKGFLLTAQLKQDGKSGRTLLALEGCLSORQFIVSNGP | 120 | | | | |
| Db | 61 | dytpvpvnaddlkitklmrqkgflltaqlkqdgksrgtllalegpgsqrfivsnngp | 120 | | | | |
| QY | 121 | ADTLDTYWDGTRHVVSLDEVDGLADSQWKNVTQVAGETYSLSHVGCDLDSFALDPPFY | 180 | | | | |
| Db | 121 | adtldtywdgtrhvvsldevdgladsqwnvtqvagetyslhvgcdllgpvaldepyf | 180 | | | | |
| QY | 181 | EHLQAEKSRMYVAKGSAHSHFRLGLQNLVHLVFNSEVEDLSKKGCGGQGAELNAISEN | 240 | | | | |
| Db | 181 | ehlqaeksrmyvakgsahshfgrllqnlvhlvfensvedlskkgcgqgggaelnaisen | 240 | | | | |
| QY | 241 | TETLRGLPHVTVTEVGPSSRRRPPVCERSCEELGNMVQELSGHLVNLQPSNLKRVSD | 300 | | | | |
| Db | 241 | tetlrghphvttvtevgpssrrppvcersceelgnmvqelsghlvnlqpsnlkrvsnd | 300 | | | | |
| QY | 301 | NQFLWELIGGPPKTRNMSACWQDGRFFAENETWVVDSCCTCTCKKFKTICHQITCPCPATC | 360 | | | | |

RESULT 3

AAB90800

ID AAB90800 standard; Protein; 1170 AA.

XX AAB90800;

XX 15-JUN-2001 (first entry)

XX Human shear stress-response protein SEQ ID NO: 100.

XX Human; shear stress-response protein; vascular disease; arteriosclerosis.

XX Homo sapiens.

OS

XX WO200125427-A1.
 XX 12-APR-2001.
 XX 02-OCT-2000; 2000WO-JP06840.
 XX 01-OCT-1999; 99JP-0280976.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA (NOJI/) NOJIMA H.
 XX Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
 PI Kuga T, Sekine S, Nakamura Y, Sugano S;
 XX WPI: 2001-266308/27.
 DR N-PSDB: AAH02923.
 XX DNA sequences, proteins encoded by them and antibodies against them
 PT useful in diagnosis and treatment of vascular disease caused by
 PT arteriosclerosis
 XX Claim 60; Page 515-521; 678pp; Japanese.
 XX The present invention provides the protein and coding sequences of a
 CC number of human shear stress response proteins. These are useful in the
 CC diagnosis, treatment and screening of vascular diseases caused by
 CC arteriosclerosis, including heart failure, post-PTCA restenosis and
 CC hypertension.
 XX Sequence 1170 AA;
 SQ

Query Match 64.0%; Score 4228; DB 22; Length 1170;
 Best Local Similarity 61.6%; Pred. No. 6.6e-253;
 Matches 725; Conservative 170; Mismatches 267; Indels 14; Gaps 6;

QY 1 MYWRL-VLLALWVWSPSTQAGHQDKD-TTFDLFSINIRKKTIGAKQFPGPPGVAYRFV 58
 DB 3 lawglvflmhvcojnripesggdsvdfelcgaarksgrrlvkqgdpaspafie 62
 QY 59 RFDYIPVWADLSTIKIMKQEGFFLTAQKDGKSGTLALLEGSLGSLORQPEIVSN 118
 DB 63 danlppvpddkfqlvdavraekflllaslrqmkktgrtllalerkdhsqgvfsvsn 122
 QY 119 GPADTLDLTYWIDTRHVVLEDVGLADSONKNTVQVAGETYSLVHVCDDLIDSPALDEP 178
 DB 123 gkagtlidslvtvgkqhvsveaalatgwkstifvgedraqlyidckmenaeldvp 182
 QY 179 F---YEHLQAEKSRMYAKGSAHSHFRGLQNVHLVPENSVEDILSKKGGQGGAGBIN 235
 DB 183 lqsvftrdlaslarlriakgvnd-nfgvlnqvrfvfgtpeidlrlnkgsstsvilt 241
 QY 236 AISENTELRGLPHVTVTYGSSRRPEVCSERCEELGNMVQELSGHLVNLQNPSENK 295
 DB 242 L--dnvnvngspairtnyghktkdlqalcgiscdelssmvlrigrirtivtlqdsir 299
 QY 296 RVSNDNQFLWELIGPPKTRNNSACWQGRFAENETWVDSCTTCCKKFTICHOJTC 355
 DB 300 kvteenkelanelirrp-----lcyhngvqyrnneetvdsctechqcnvltckkvc 353
 QY 356 PPATCASPSFVGECCFSLHSDVEEGSWPAEWTCQSVTCGGTQQRGRSCDVTNVC 415
 DB 354 plmpcsnatvpgccprcwpsdsaddgswpsewtscstscnglqgrfscdslnnrc 413
 QY 416 LGPSTQTRACLSKCDTRIQDGGHSHSPWSSCSVTCGVGNITRILCNBPVPMGKN 475
 DB 414 egssvqtrtchicqdkrfkqggghwspsscsvtcgvitrlcnspspqmgkp 473
 QY 476 CKGSGRETKACOGAPCPIDGRWSPWSACTVTCAGGIRETRVCNSPEQYGGKACVG 535
 DB 474 cegaretkackackcpingvgwpswdlscvtcggvqqrksrlnnppapqfggkdcvg 533

QY 536 DVOERQMCNKRSRCPVDGCLSNPCRPAGOCSSPFDGSGSCGSPVGLNGTHCEDLDECA 595
 DB 534 dvtengicnkqdcpldgclsnpcrfagvictsytpdsgswkcgacppgysnglqctadvack 593
 QY 596 LVPDIFCSTSKVPRCVNTPQGFHCLPCPPRYRGNOVPVGVLEAAKTEKOVCEPENCKDK 655
 DB 594 evpdacfnhgehrcentdpynclpcprftgsqpfggvehatankqvckprpctdg 653
 QY 656 THNCHKHAEICYLGHFSDPMYKCECOTYAGDGLLTCGSDSLDGPNLNLVCATNATVHC 715
 DB 654 thdcnkakcnlyghysdpmrceckpgyagngilicgedtdldgwpnenlvcvnanatync 713
 QY 716 IKDNCPLHNSQEPDFDKDGIQDADDDNDNDGVVDEKDNQCOLLENPRQADYDKDEVGDR 775
 DB 714 kkdncplnpsqgedykdigdadcdndkdpdrncpfnypaqdydrddvgr 773
 QY 776 CQNCPIVHNPAQIDTDNNGEGDACSVDIDGDDVFNERONCPIVYNTDQDRTDGDGVGDHC 835
 DB 774 cdncpynhpqadtndnngedacaadldgdilnerdncgyvynvgrtdmdvgdgc 833
 QY 836 DNCPLVHNPQDQTDVNDLVGQDQNNEDIDDDGHONNODNCPYISNANOADHRRDQGDGA 895
 DB 834 dncplehnpdqldsdridgtcdmngqldedghnmlndncpyvnanqadhdkgkda 893
 QY 896 CDPDNDNDGVDDRNCLRVNPDQEDLDGDRGCDICKDDFDNDNIPDIDDVCPENNAIS 955
 DB 894 cdhddndgipdkdncrlvpnpdqkdsdgdgrgackddfdhdsvpdliddicpenvids 953
 QY 956 ETDFRNFQWPLDPKGTQIDPNWVIRHQKELVQANSDFGIAVGFDFEGSVDFSGTFY 1015
 DB 954 etdfrfqlmldpdkgtsgndpnwvrhqqkelvtvncdpglavgydefnavdfsgtff 1013
 QY 1016 VNTDRDDYAGVFGYQSSRFYVVMKQVTQTYWEDQTRAYGYSGLKVVNSTTGTG 1075
 DB 1014 interddayagfvfygqssrfyvmwkvqtsywdtnptraqgysglsvkvnsttggp 1073
 QY 1076 EHLRNALWHTGNTPGQVRLMHPDNICGWKDYATAYRWHLTHRPKTYIRLVVHEGQVMA 1135
 DB 1074 ehlnalwhtgntpgqvrtlwhdprhgwkdftayrwlshrpktgfrvrmvgegkima 1133
 QY 1136 DSGPIYDQTYAGRGLGVFQSEMYVFSDLKVECD 1171
 DB 1134 dsdpiydktyaggrlglvfseqemvfdsdkyecd 1169

RESULT 4
 AAB74450
 ID AAB74450 standard; Protein; 1170 AA.
 XX AC AAB74450;
 XX DT 06-JUN-2001 (first entry)
 XX DE Human variant thrombospondin 1.
 XX KW Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
 KW polymorphism; vascular disease; coronary artery disease; forensics;
 KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
 KW pulmonary embolism; paternity test.
 XX OS Homo sapiens.
 XX PN WO200118250-A2.
 XX PD 15-MAR-2001.
 XX PF 07-SEP-2000; 2000WO-US24503.
 XX PR 10-SEP-1999; 99US-0153357.
 PR 26-JUL-2000; 2000US-0220947.
 PR 16-AUG-2000; 2000US-0225724.
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES. FA

PA (MILL-) MILLENNIUM PHARM INC.
XX Lander ES, Gargill M, Ireland JS, Bolk S, Daley GO, McCarthy JJ;
XX N-PSDB; AAF95238.
XX WPI: 2001-226749/23.
XX DR N-PSDB; AAF95238.
XX Nucleic acids comprising single nucleotide polymorphisms, useful in
XX applications such as forensics, paternity testing, medicine, genetic
XX analysis and phenotype correlations to diseases such as diabetes and
XX atherosclerosis -
XX Claim 20; Fig 1; 242pp; English.
XX The present invention provides a method of diagnosing a vascular disease
XX in an individual, involving determining the sequence at various
XX polymorphic sites within the human thrombospondin 1 and thrombospondin 4
XX genes. The sequences at a number of polymorphic sites are also provided
XX in the specification. In particular, the method can be used in the
XX diagnosis of atherosclerosis, myocardial infarction, coronary heart
XX disease, stroke, peripheral vascular diseases, venous thromboembolism
XX and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also
XX useful in forensics, paternity testing, genetic analysis and phenotype
XX correlations to diseases. The present sequence is the human variant
XX thrombospondin 1 protein.
XX Sequence 1170 AA;
XX
Query Match 54.0%; Score 4228; DB 22; Length 1170;
Best Local Similarity 51.6%; Pred. No. 6.6e-253; Indels 14; Gaps 6;
Matches 725; Conservative 170; Mismatches 267;
QY 1 MYWRL-VLLALWVSPSTQAGHDKD-TTFDLFSISNINRKTICAKQFGPDPGVPAFYRFV 58
Db 3 lawglvflmhcvtgnripesggnsvdfelftgaaarksgrrllvkgpdpapafrie 62
QY 59 RFDYTPPVNADLSKITIMROKEFFLTAQLKQDGKSGRTLLALEGPGLSORQFEIVSN 118
Db 63 danllppvddkfqldvavraekflllaslrqmkktgrtllalerkdhsqgvfsvsn 122
QY 119 GPADTLTYIDGTRHYVLESDVGLADSONKNVQVAGETYSLHVGCGLDLSDPALDEP 178
Db 123 gkagtlidslvtgqkqhvsvaeallatqgwkstlfvqdradqlvdeekmenaelvdp 182
QY 179 F---YEHQAERKMYVAKGSAESHFRGLLQNVHLVFNSEVEDILSKKGGCQGGAGAEIN 235
Db 183 lqsvftrdlaslarlriakgvnd-nfgvqlqnvrfvgttpeulrnkgsstsvilt 241
QY 236 AISENTEILRLGPHVTETVGVSSRRPEVCERSCEELGNMVOELSLHLVLPQSENLK 295
Db 242 l--dnvnvngspairtnyghktldlqalcglsdelssmvlairglrtivttlqdsir 299
QY 296 RVSNDNQFLWELIGPPKTRNNSACWQDGRFAENETWVVDSCCTCTCKKFTICHQITC 355
Db 300 kvtcenkelanelrtp-----lcyhngvqyrneetvdsctechcqnsvtckkvc 353
QY 356 PPATCASPSFVEGECPCFLSHVDGEGSPWAEWTQCSVTGCGTQQRGSCDVTNFC 415
Db 354 plmpcsnatvpdgeccprcwpsadsaddgswpsewtscstscnglqgrscdslnnrc 413
QY 416 LGPSTQTRACSLSKCDTRTRQDGGSHSWSPSSCSVTGCGVGNITRILCNPSVPQMGKN 475
Db 414 egssvqtrtchigecdkrfkqdgghswspsscsvtcgvgvitrilcnpspdmngkp 473
QY 476 CKSGRETKACGACPCFIDGRSPSPSACTVTTCAGGIRERTRVCNSPEPQYGGKACV 535
Db 474 cegearctackdcpingvgwpswdicvctgvgvqkrslcnnpapafgkdcvg 533
QY 536 DYQERQMKRSCVPDVCGLSNPCFFPGAQCSFPDGSWSGCPVGFGLNGTHCEDLDECA 595
Db 534 dvtenqicnkqcdpdcgpcagvkcstypdgswwkcgacppgysnglqctdvdeck 593

QY 596 LVPDICFSTSKVPRCVNTOPGFHCLPCPPRYRGNOPVGVGLEAAKTEKOVCSBPENCKOK 655
Db 594 evpdacfnhnghrceatdpgvncipcpbrftgspfggvehatankqvcxprnptcdg 653
QY 656 THNCHKHAEIYLGHFSDPMYKCEQOTGYAGDGLICGEDSLDGPENLNLVCATNATYHC 715
Db 654 thdcnkakcnlylghysdpmrceckpgyagngilcgedtdldgwpnenlvcvanatyhc 713
QY 716 IKDNCPLHNSQOEDEFDKDIGDAGDDDDNDGVTDEKNCQLLENPRQADYDKDEVGDR 775
Db 714 kkdncpnlpnsgeydkdgidgacddndkdpdrdncpfhynpacydydrdvgdr 773
QY 776 CDNCPIYVHNPAQIDTNDNGEGDACSVDIDGDDVFNERDNCPIYVNTDQDRTDGDGVDHC 835
Db 774 cdncpynhnpdadtndngegdaadidgdgilnerdncgyvynvqgrtdmdgvgdgc 833
QY 836 DNCPLVHNPQDQVDNDLVGDDCDDNEDIDDBGHQNNQNCPIYSANQADHRRDQGDGA 895
Db 834 dncpllehnppdqldsdgsdrigtdcdmndldedghqnnldncpyvpnanadhdkgkda 893
QY 896 CDPDNDNDGVPDRDNCRLVFNPDQEDLDGDRGDRICKDDFDNDNIPDIDDDVCPENNAIS 955
Db 894 cdhddndngipdkdncrlvnpdpckdsdgdrgdackdhdsvpdldiddicpenvdis 953
QY 956 ETDFRNFQWPLDPKGTQIDPNWIRHCKELVQFANSDDPGIANGFDFSGVDFSGTFY 1015
Db 954 etdfrfmgipldpkgtsqndpnwvrrhkgkelvqtnvcdpglavgydefnadvstgftf 1013
QY 1016 VNTDRDDYAGVFYQSSRFYVVMKQVOTYTWEDQPTRAYGYSGLKVVNSTGTG 1075
Db 1014 interddayagfvygqssrfyvmwkvqtsywdtpttraagysglskvvnstgpg 1073
QY 1076 EHLRNALMHTNTPGCVRTLWHDPRNIGWKDYATAYRWHLTHRPKTYGIRLVHVEGQVMA 1135
Db 1074 ehlrnalwhtnpgqvtlwhdprhgwkdftayrwlshrpktgfrvrmvvegkima 1133
QY 1136 DSGPIYDOTYAGRGLGFVFSOEMVYFSDLKVECDR 1171
Db 1134 degpiydktyagrglgrlgrfvsqemvffsdlkyecrd 1169
RESULT 5
AAB00042
ID AAB00042 standard; Protein; 1152 AA.
XX AAB00042;
XX 08-NOV-2000 (first entry)
XX Human thrombospondin-1 (TSP-1).
XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
XX thrombospondin; angiogenesis; tumour; treatment; cancer;
XX arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
XX glaucoma.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Region 361..416
XX /label= Type 1 repeat region
XX Region 417..473
XX /label= Type 1 repeat region
XX Region 474..530
XX /label= Type 1 repeat region
XX WO2000044908-A2.
XX 03-AUG-2000.
XX 01-FEB-2000; 2000MO-US02482.
XX 01-FEB-1999; 99US-0118053.
PR

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PI Lawler JW;
 XX WPI: 2000-514823/46.
 DR Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 XX matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 PT Disclosure; Fig 1; 40pp; English.
 XX
 XX New nucleic acids are described which encode a protein comprising
 CC the second and third type-1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX Sequence 1152 AA;
 SQ

Query Match 64.0%; Score 4226; DB 21; Length 1152;
 Best Local Similarity 62.5%; Pred. No. 8.6e-253;
 Matches 719; Conservative 167; Mismatches 253; Indels 12; Gaps 4;

Qy 24 DTTFDLFSINIRKRTIGAKFRGPDGPVAYFRVFDYIPVNNADLSKTKIMROREQ 83
 Db 10 nsavdiftelgaarkgsgrlvgkdpaspafriedanlipvpddkfgdlvdartekg 69
 Qy 84 PFLTAQLKQDKSRGTLALALEGLSQRFELVSNQFADTLTYWIDGTRHVSLEVDG 143
 Db 70 fillaslrqmktrgillalerkdhsqgfvsvngskagtdlslvtvgkghvsveaal 129
 Qy 144 LADSOAKNVTVOVAGETVSLRVGCDLIDSEALDEPF---YEHLOAKSRMVVAKGSARES 200
 Db 130 latgqwkstifvqdraglyldeekmenaeldvpqlqsvfrdlasiariakggvnd- 188
 Qy 201 HFRGLLQNVHLFVENSVEDILSKGCGQOGAEINAISENTETTLRLGPHVTEYVGPSE 260
 Db 189 nfgvqlnvrfvgtptedilrnkqcsstsvlltl--dnvngvsspairtnyghktk 246
 Qy 261 RPEVCESCEELGNWVQELSLHVLVQNPENLKRVSNDNQFLWELIGGPPKTRNMSAC 320
 Db 247 dlqalcigscdeislmvlelrglrvltltdlsirkvteenkelanelrrpp-----lc 300
 Qy 321 WODGRFAENETWYVDSCTTCTCKFKTKITCQITCPATACSPSFVEGECPCSLHVSVDG 380
 Db 301 yhngvqyrneewtvdstechnqsvtlckvscplmpcsnatvdpgeccprcpsda 360
 Qy 381 BEGNSPWAETWOCSTVTCGSGTQGRSDVTNTCLGPSIQTRACSLSKCDTRINQDGGW 440
 Db 361 ddgswpsweatscscnglqgrgrscdslnnrcegsqvtrtchiqecdkrfkqdgw 420
 Qy 441 SHWSPWSSCVTCGVGNTRILCNLSPVPMQGNKCKSGRETKACQACAPCPIDGRNSPW 500
 Db 421 shwspwsscvtcgvgvtrilcnlspvpmqngkpcgearetkackkdacpingswgpw 480
 Qy 501 SPWSACTVTCAGIRTRVNCSPPOYGGKACVGDVQERQCNKRKSCVPDGLSNPCFP 560
 Db 481 spwdicstcvgvgvqkrarlcnnptpqfggkdcvgdvntgncnkqdcpiqclsnpcfa 540

Qy 561 GAQCSFSDGSMSCGSPVGLNGTHCEDLDECALVPDICTSTKVPKPCVNTQGFHCL 620
 Db 541 gvktsypdgswkcgacppgysnglqctdvdcekevpadcfhnhgehrcentdpgyncl 600
 Qy 621 PCPPRYRGNQPVGVGLEAAKTEKQCEPENPCDKTHNCHKAECIYLGHSFSDPMYKCEC 680
 Db 601 pcpprtfsgpfqgvehatankgvckprnctdgtchcnknaknynlghysdpmvrec 660
 Qy 681 QTGAGDGLICGEDSLDQGNLNLVCATNATYHICIKONCPHLNPSGGQEDFKDGIQDAC 740
 Db 661 kpgyagngllicgedtldgwpnenlvcvanatyhckdncnplnpsggedydkdglgac 720
 Qy 741 DDDDDNDGVTDEKNCQQLLFNPRQADYDKDEVGRDCDNCPCYVHNPQAQIDTDNNGEGDACS 800
 Db 721 ddddddnhkdpdrncpfnypaqdydrddvgrdcncpynhnpdqadtdnnggedaca 780
 Qy 801 VDIDGDDVFNERNDCPYVYNTDQRTDGGVGDHCDNCPLVHNPDQTDVDNDLVGQDCDN 860
 Db 781 adidgdgillnerdncqyvnyvqrdtdmdvgdgcncplehnpqlddsdrlgdtcdn 840
 Qy 861 NEDIDDDGHONQDNCPCYISNANQADHDDGGDGDACDDDDNDGVPDDRDNCRLVFNPDQ 920
 Db 841 nqdidgdhqnldncpypvnanagadhkdkgkgsdchddndgipddkdnclrvnppdq 900
 Qy 921 EDLDGDRGDIKDDFDNDNIPDIDVCPENNAISETDFRNFQMVPLDPKGTQIDPNMY 980
 Db 901 kdsdgdrgdackddfdhdsvpiddicpenvidsetfrfmgpldpkqtsqndpnwv 960
 Qy 981 IRHQGKELVQANSDPGIAVGFDFGVSFGTFFVYNTDRDDDYAGFVFGVQSSSRFYV 1040
 Db 961 vrhgkkelvqtvncdpglavgydefnavdfsgtffinterdddyagfvgysssrfyv 1020
 Qy 1041 MWKQVTOYWEQDQTPRAYGYSVSLKLVNNTGTGHEHLRNALWHTGNTPGQVRTLWHDPR 1100
 Db 1021 mwkvqtsywdntntraggysglsvkvvnstgpghehlrnalwhtcgnlcpvgvrtlwhtpr 1080
 Qy 1101 NIGWKDYATYRHLTHRPKTYIRLVHREGQVWADSGPIYDQTYAGGRLGLFVFSQEMV 1160
 Db 1081 hlgwkdfctay-wrlshrpktgflrvmyegkkmadsqlydktyaggrlgflvfseqmv 1140
 Qy 1161 YFSDLKYBCRD 1171
 Db 1141 ffsdkkyecrd 1151

RESULT 6
 AAG75130
 ID AAG75130 standard; Protein; 571 AA.
 XX
 AC AAG75130;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:5894.
 DE
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 6.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-235357/24.
DR N-PSDB; AAH34535.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7409-7411; 9803pp; English.
XX AAH32943 to AAH37195 and AAH37514 to AAH77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps;
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX Sequence 571 AA;

Query Match 49.2%; Score 3251; DB 22; Length 571;
Best Local Similarity 100.0%; Pred. No. 6.8e-193;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 SKVPRCVNTQPGFHCIPCPRRYGNQPVGVGLEAAKTEKQVCEPENPCPKDKTHNCHKHAE 664
DB 4 skvprcvntqpgfchelpcrrygnqpvvgvgleaaktekqvcpenpcdkthnchkhae 63
QY 665 CIYLGHFSDPMKCEQCTQYAGDGLICGEDSLDGNPNLNCATNATYHCIKDNCPLHP 724
DB 64 ciylghfsdpmkcectqyagdglicgedslldgpnlnlvcnatnatyhcikdncplhp 123
QY 725 NSGOEDFKDGTGDCADDDDDNGVYDEKDNLCOLLFPQADYDKDEVDRCNCPYVHN 784
DB 124 nsqgedfkdgtdgacdddddngvtdkdnclcollfprqadydkdevdrcncpyvhn 183
QY 785 PAQIDTNNNGEGDACSVDIDGDDVFNERNDCPVYNTDQDRTDGGVGDHDCNCPVHNHP 844
DB 184 paqidtnngedacsvdldgddvfnerndcpvyntdqdrtdggvgdhdncncplvhn 243
QY 845 DQFDVNDLVGQCDNNEDIDDGHQNNQDNCPIYSNANQADHRDGGDAGCDPDDNDG 904
DB 244 dqfdvndlvqcdnnediddgghqnnqdnncpiysnanaqadhrdggdgcddpddndg 303
QY 905 VPDRNCRNVNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISDTERNFOM 964
DB 304 vpdrnrcnlvnpdqedldggrgdickddfdndnipdidvcpennaisdternfom 363
QY 965 VPLDPKGTQIDPNMWRHQQGKELVQANSDPGIAVGDFEFGSVDFSGFYVNTDRDDY 1024
DB 364 vpldpkgtqidpnwvrrhqqgkqlvqansdpgiavgdfefgsdfgyvntdrdday 423
QY 1025 AGFVFGYQSSRFYVVMKQVQTQYWEQDPTRAYGVSGLKVVNSSTTGTGHELRNALWH 1084
DB 424 agfvfyqssrfyvwmkqvqtqyweqdptraygysglkvvnsttgtgehlrnalwh 483
QY 1085 TGNTPQCVTLWHDPRNIGKDYATYRWHLTHRPKTYIRLVLHVEGKQVMADSGPIYDOT 1144
DB 484 tgntpqcvrtlwhdprnigkdyatyrwhlthrpktgyirvlvhegkqvmadsgpiydot 543
QY 1145 YAGGRGLGVFVSQEMVYFSDLYVECRDI 1172

DB 544 YAGGRGLGVFVSQEMVYFSDLYVECRDI 571
RESULT 7
AAB43586
ID AAB43586 standard; Protein; 571 AA.
XX AAB43586;
AC AAB43586;
XX 08-FEB-2001 (first entry)
XX Human cancer associated protein sequence SEQ ID NO:1031.
KW Human; cancer associated gene; cancer antigen; detection; cancer;
diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
antiflammatory; antithyroid; antiallergic; antibacterial; cardiac;
dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;
vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
immune disorder; haematopoietic cell disorder; autoimmune disorder;
allergic reaction; graft versus host disease; organ rejection;
haemostatic; thrombolytic; cardiovascular disorder; infection;
neurological disease; drug screening.
XX Homo sapiens.
XX WO200055350-A1.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05882.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2000-587533/55.
XX N-PSDB; AAC77795.
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX Claim 11; Page 1616-1619; 2352pp; English.
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
in AAB43398 to AAB44239. The proteins can have activities based on the
tissues and cells the genes are expressed in. Example of activities
include: cytostatic; proliferative; vulnerable; immunomodulator;
antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
antiflammatory; antithyroid; antiallergic; antibacterial; antiviral;
dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
nocotropic; vasotropic; antipsoriatic and antiangiogenic. The
polynucleotides and polypeptides can be used for preventing, treating or
ameliorating medical conditions and diagnosing pathological conditions.
CC polynucleotides, antibodies, agonists and antagonists from
the present invention may be used to treat immune disorders by activating
or inhibiting the proliferation, differentiation or mobilisation of
immune cells, to treat disorders of haematopoietic cells, autoimmune
disorders, allergic reactions, graft versus host disease and organ
rejection, modulate haemostatic or thrombolytic activity, modulate
inflammation, cancers, cardiovascular disorders, neurological disease and
bacterial or viral infections. The peptides, nucleotides, antibodies,
agonists and antagonists may be also be used in drug screens. AAC78449 to
AAC78457 and AAB44240 represent sequences used in the exemplification of
the present invention.
XX Sequence 571 AA;

Query Match 49.0%; Score 3236; DB 21; Length 571;

| | | | |
|--|--|--|--|
| Best Local Similarity 99.5%; Pred. No. 5.7e-192; | | Matches 565; Conservative 0; Mismatches 3; Indels 0; Gaps 0; | |
| QY | 605 SKVPCVNTQPGFHCILPCPPRYRGNOVGVGLEAAKTEKQVCEPENPCPKDKTHNCHKHAE 664 | | |
| Db | 4 skvpcvntqpgfhclpcppryrgnpgvgleaaktekqxcpenpcdkthnchkhae 63 | | |
| QY | 665 CIYLGHFSDPMYKCEQOTGAGDGLICGSDSDLDGHPNINLVCAATNATYHCIKDNCPHLP 724 | | |
| Db | 64 ciylghfspmkykceqotgagdglicgsdldgwpnlnlvcatnatyhcikdncphlp 123 | | |
| QY | 725 NSQOEDFDKDGIGDADDDDDNDVTDKDNQCLLFNPRQADYDKDEVGDRCDNCPYVHN 784 | | |
| Db | 124 nsqoedfdkdgigdadddndvtdkdnqcllfnpqadydkdevgdrdcncpyvhn 183 | | |
| QY | 785 PAQIDFNNNGEGDACSVDIDGDVFNERNDCPYVNTDQDITDGDGVDGHCNCPLVHNP 844 | | |
| Db | 184 paqidfnnngedgacsvdidgdvfnerncnyvntdqdrtgdgvgdhcncplvnhp 243 | | |
| QY | 845 DQTDVDNDLVGQCONNEDIDGHHONNODNCPYISNANOADHRRDGGDADCDPDDNDG 904 | | |
| Db | 244 dqtdvdndlvqgqcnndedidgghonndncpyisnanaqdhrrdggdadcddndg 303 | | |
| QY | 905 VPDDRNCRLVFNPOEDLDGSGRGDICKDDFDNDNIPDIDVCPENNAISSETDFRNFOM 964 | | |
| Db | 304 vpdtrncrlvfnpqedldgsgrgdickddfdndnlpaidvcpennaisetdfnfm 363 | | |
| QY | 965 VLDPKGTQIDPNVIRHOGKELVOTANSDPGIAVDFDFGVSDFGFFVNTDRDDY 1024 | | |
| Db | 364 vldpkgtqidpnvvrhkgelvtansdpgiavdfdfgvsdfgffvntdrddy 423 | | |
| QY | 1025 AGFVFGYQSSRFYVVMKQVNTQVWEDQPTRAYGSGVSLKVNSTGCTGHLNALWH 1084 | | |
| Db | 424 agfvfyqssrfyvwmkqvntqvwwedqptraygsgvslkvnstgtgghlnalwh 483 | | |
| QY | 1085 TGNTPQVRLTWHDPNIGWKDYATYRWHLTRPKTYIRLVHLEGKQVMDSGPIYDQT 1144 | | |
| Db | 484 xgntpgvrltwhdpnigwkdyatyrwhltrpktgyirvlvhegkqvmadsgpiydt 543 | | |
| QY | 1145 YAGGRGLFLVFSQEMVYFSDLKYECRDI 1172 | | |
| Db | 544 yaggrglflvfsqemvyfslkyecrdi 571 | | |
| RESULT 8 | | | |
| ID | AAU02913 | | |
| XX | AAU02913 standard; Protein; 731 AA. | | |
| AC | AAU02913; | | |
| XX | 12-SEP-2001 (first entry) | | |
| DE | Angiotensin converting enzyme (ACEV) splice variant, protein #13. | | |
| XX | Angiotensin converting enzyme splice variant; ACEV; interleukin 6; | | |
| KW | granulocyte colony stimulating factor receptor; glucagon; hypertrophy; | | |
| KW | platelet-derived endothelial cell growth factor; cardiovascular disease; | | |
| KW | cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C; | | |
| KW | vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer; | | |
| KW | myocardial infarction; coronary arterial thrombosis; renal disease; | | |
| KW | diabetic nephropathy; muscular disease; immune disorder; sarcoidosis; | | |
| KW | multiple sclerosis; immune complex nephritis; deep vein thrombosis; | | |
| KW | nonaroidotic pulmonary granulomatous disease; endothelial abnormality; | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | WO200136632-A2. | | |
| XX | | | |
| PD | 25-MAY-2001. | | |
| XX | | | |
| PF | 17-NOV-2000; 2000WO-IL00766. | | |
| XX | | | |

| | | | |
|----|---|--|--|
| PR | 17-NOV-1999; 99IL-0132978. | | |
| PR | 10-DEC-1999; 99IL-0133455. | | |
| XX | (COMP-) COMPUGEN LTD. | | |
| PI | Levine Z, David A, Azar I, Khosravi R, Bernstein J; | | |
| XX | WPI; 2001-336004/35. | | |
| DR | N-PSDB; AAS06013. | | |
| XX | Novel alternative splicing variants e.g. variant of angiotensin | | |
| PT | converting enzyme (ACEV), useful in identifying candidate compounds | | |
| PT | capable of binding to the variant and to detect anti-variant antibodies | | |
| XX | Claim 4; Fig 13; 519pp; English. | | |
| XX | The sequence represents an angiotensin converting enzyme splice variant | | |
| CC | (ACEV) polypeptide. The polypeptides of the invention include variants of | | |
| CC | granulocyte colony stimulating factor receptor, glucagon, interleukin 6, | | |
| CC | platelet-derived endothelial cell growth factor, cyclin-dependent kinase | | |
| CC | inhibitor 1C, cellular tumour antigen P53, and vasoactive intestinal | | |
| CC | polypeptide receptor 2. The polypeptides and their associated nucleic | | |
| CC | acids are useful for identification of variant sequences and detection of | | |
| CC | candidate compounds capable of binding to the molecules. The sequences of | | |
| CC | the invention can be used in the treatment and diagnosis of various | | |
| CC | disorders including cardiovascular diseases such as arteriosclerosis, | | |
| CC | myocardial infarction and coronary arterial thrombosis, renal diseases | | |
| CC | such as diabetic nephropathy, muscular diseases such as hypertrophy, | | |
| CC | immune disorders such as immune complex nephritis, multiple sclerosis, | | |
| CC | cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such | | |
| CC | as asbestosis and vascular pathologies involving an endothelial | | |
| CC | abnormality such as deep vein thrombosis. | | |
| XX | Sequence 731 AA; | | |
| QY | Query Match 31.7%; Score 2094; DB 22; Length 731; | | |
| Db | Best Local Similarity 50.1%; Pred. No. 3e-121; | | |
| Db | Matches 365; Conservative 123; Mismatches 227; Indels 14; Gaps 6; | | |
| QY | 1 MWRL-VLLALWMPSTQAGHQDKD-TTFDLFSINIRKKTIGAKQFQPDGVPAYRFV 58 | | |
| Db | 3 lawglvflmhvqctnrtpesgdnsvdfidelfelgaarksgrrlvkgpdpssapafrie 62 | | |
| QY | 59 RFDYIPPNADDLKSKITIMRQKEGFFTAQLKQDGKSGRTLLALEGPGLSQRQFEIVSN 118 | | |
| Db | 63 danlippvddkfgdlvdavraekgflillaslrqmktrgtllalerkdhsgqvsvsn 122 | | |
| QY | 119 GPADTLDLTYYWIDGPRHVVSLEDDVGLADSONKNTVOVAGETYSLVHVGCDLDSFALDEP 178 | | |
| Db | 123 gkagtdlsltlvqgkghvsvveeallatcgwksitlfvqedraqlydcekmenseldvp 182 | | |
| QY | 179 F---YEHQAERKSMYVARGSARESHFRGLLQNLVHNFVENSVEDILSKKGGQGGAEIN 235 | | |
| Db | 183 lqsvtrdlasrlrlakgvnd-nfgvlgvqnrvfvtgtpedilrnkgsstsvilt 241 | | |
| QY | 236 AISENTELRGLPHVTTFYVGPSSERRPEVCSERCEELGNMVVQELSGHLVNLQNPSENLK 295 | | |
| Db | 242 l--dnvvngvsgspairtnyghktldlqalcgicdelssmvlrgrlrvttlqda 299 | | |
| QY | 296 RVSDNDFLWELIGPPKTRNMSACWODGRFFAENETVWVDSCTTCTCKKFKTICHOITC 355 | | |
| Db | 300 kvteenkelanelrrpp-----lcyhngvgyrnnnevtvdsctechqnsvtickkvsc 353 | | |
| QY | 356 PPATCASFVEGECPCSLHSVDGEGSWPAEWTCQSVTCGSGTQOGRSCDVTNVC 415 | | |
| Db | 354 pmpcsnatvpdgeccprcwpsdsaddgswpsewtscstcngiqgrgrscdlnnrc 413 | | |
| QY | 416 LPSIQTRACSLSKCDTRIRDGGWSHSPWSSSVTCGVCNITRILCNFPVQMGNK 475 | | |
| Db | 414 egssvqtrtchicgckrfkqdgwshwspwsscsvtcgdvitrilcnspqmgk 473 | | |

AA093335;
06-NOV-2001 (first entry)
Human polypeptide, SEQ ID NO: 2869.
Human; full length cDNA; cDNA synthesis; oligo-capping.
Homo sapiens.
EP1130094-A2.
05-SEP-2001.
07-JUL-2000; 2000EP-0114089.
08-JUL-1999; 99JP-0194496.
11-JAN-2000; 2000JP-0118774.
02-MAY-2000; 2000JP-0183765.
(HELI-) HELIX RES INST.
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
WPI; 2001-524255/58.
N-PSDB; AAK94255.
830 Primers useful for synthesizing full length cDNA clones and their
use in genetic manipulation -
Claim 8; SEQ ID NO 2869; 1380pp + sequence listing; English.
The invention relates to primers for synthesizing full length cDNA
clones. 830 cDNA molecules encoding a human protein have been
isolated and nucleotide sequences of 5' and 3' ends of the cDNA
molecules have been determined. Primers for synthesizing the full length
cDNA are useful for clarifying the function of the protein encoded by
the cDNA. The full length clones were obtained by construction of full
length enriched cDNA libraries that were synthesised by the oligo-capping
method. The primers enable the production of the full length cDNA easily
without any special methods. The present sequence is a polypeptide
encoded by a full length human cDNA of the invention.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in CD-ROM format directly from EPO.
Sequence 762 AA;
Query Match 29.58; Score 1946.5; DB 22; Length 762;
Best Local Similarity 52.28; Pred. No. 4.1e-112;
Matches 360; Conservative 78; Mismatches 183; Indels 69; Gaps 11;
QY 537 VOEROMC---NKRSCPVDGCLSNPCFCAQCS---SPFDGSMSCGCPVGLNGTHCED 590
DB 69 imecvcgfheirs-----hcsnpfcrgvdcmeveyep--gyrcgpcpglqngthcsd 122
QY 591 LDECALVPDIFSTSKVPRCVNTOPGFHCLPCPPRYRGNQPVGVGLEAAKTEKQVCE--- 647
DB 123 inecaha-dpcfpss---clntmpgfceacprgkgtqvgvgidyaraskvcndid 178
QY 648 -----PENPCDKDTIN-CHKHAEC 665
DB 179 ecndngnggcpnsictntvsgfkcgprlglfnqsgcglpartchspahspchihaic 238
QY 666 IYLGHFSPMYKCEQOTYAGDGLICGEDSLDGLWPNLNLVCATNATVHC1KDCNCPHLPN 725
DB 239 if---erngavscnqvgwagncvgtcdtdidydpdqlpcmdn-nkhckqdcncltbn 294
QY 726 SQGEDFDKDGICDADDDDDNDVTDKDNQQLFNPRQADYDKDEVDGDCNCPYVHNP 785
DB 295 sggedadndvgdcdaddadgiknvedncrlfpnkdaqnsdtdsfgdacnncpnvnn 354

QY 786 AQIDTNNEGDACSVDIDGDDVFNERNCNPVYVNTDQDTDGCGVGHDCNCPVLVHNP 845
DB 355 dqktdngsgedacndndvsgdipnglancpkvpnpqlctdrdedyvgdacsccpemsapt 414
QY 846 QTDVNDLVGDCDNNEDIDDDHONNODNCPYISNANQADHRDGGQDADCPDDNDGV 905
DB 415 qtdadsdlvgdcdtnedsdgdgtdkncpcqpinssqldsdndglgdecddndgi 474
QY 906 PD---DRDNCRLVFNPDQEDLDGGRDICKDDFDNDNIPDIDVCPENNAISTDFRN 961
DB 475 pdyvpdpdncrlvpnpnqkdsdngvgvceddfndavdvdpdvpcpsaeavtldfra 534
QY 962 FQVPLDPKGTQIDPNMVIHQKELVQTANSPDGIAGVDFEGSVDVSGTFYVNTDRD 1021
DB 535 yqtvvldpegdaqidpnwvlnqgmewtmsdpglavgytafngvdfegtfhvntvtd 594
QY 1022 DDYAGVFGYSSSRFFVVMKQVQTYWEDQPRAYGYSGLKVVNSTTGTGHEHLNA 1081
DB 595 ddyagflfsyqdsgrfvyvmwktqetwqatpfravaqplqlkavtsvsgpgehlrna 654
QY 1082 LWHGTGNTPGQVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYIRVLVHKGQVMADSGPIY 1141
DB 655 lwhghtpdqvrllwtddprnvgwrktsyrwqlhrrpqvgvyrklyegpqlvadsgvll 714
QY 1142 DQTYAGGRLGLFVFSQEMVYFSDLKYECRD 1171
DB 715 dtsmrgrglgvfcfsqenilwslqyrcnd 744
RESULT 11
AAR56248
ID AAR56248 standard; Protein; 889 AA.
XX AAR56248;
XX AAR56248;
DT 20-JAN-1995 (first entry)
XX Xenopus thrombospondin-4.
XX Xenopus thrombospondin-4.
XX Thrombospondin-4; transgenic animal.
XX Xenopus laevis.
XX WO9413794-A.
XX 23-JUN-1994.
XX 03-DEC-1993; 93WO-US11725.
XX 04-DEC-1992; 92US-0985296.
XX (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX Lawler JW;
XX WPI; 1994-217874/26.
DR N-PSDB; AAK66454.
XX Human thrombospondin-4 - made by recombinant DNA methods, useful
PT in experimental and therapeutic protocols
XX Disclosure; Page 45-47; 66pp; English.
XX cDNA encoding X. laevis thrombospondin-4 was cloned by PCR using
CC degenerate primers (AAQ66456-57) based on conserved regions of various
CC thrombospondins. The cDNA sequence (AAQ66454) and corresponding amino
CC acid sequence (AAR56248) were determined.
XX Sequence 889 AA;
Query Match 28.88; Score 1899; DB 15; Length 889;
Best Local Similarity 51.98; Pred. No. 4.2e-109;

Db 934 snlkyrcnd 942
RESULT 13
ABB65649
ID ABB65649 standard; Protein; 1024 AA.
XX AC ABB65649;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 23739.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WIPI; 2001-656860/75.
XX DR N-PSDB; ABL09752.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS Disclosure; SEQ ID NO 23739; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1024 AA;
Query Match 27.3%; Score 1805; DB 22; Length 1024;
Best Local Similarity 34.0%; Pred. No. 3,2e-103;
Matches 407; Conservative 134; Mismatches 354; Indels 302; Gaps 31;
Qy 77 IMRQKEGFLTAQKQDGKSGTLLALEGPGLSQ-ROFEIVSNGPADTLDLYWIDTRH 135
Db 6 mlkhmsfll-----drkqrvldisangatesrnfelplnetati----- 48
Qy 136 VVSLVDGLADSQMKNVTVOAGETYSLVHVCGLDLSFDALDEPFYHLOAEKSRMYAVK 195
Db 49 -----rstaigfskrithlvdcaksthhdid-----mmlakiyqmd 86
Qy 196 SARESHFRGLQLQVHLFENSVEDILSKGQQG---QGAENAISENTEFRLRGPVH-- 250
Db 87 dpviklfrerkyph--fdgmehslqancqkgnhrgnrmknkitergeifpirlg 144
Qy 251 -----TTEV-----GPSSRRPEV---CER----- 268

Db 145 ylpqymiyisawypifaeaknkrdvrgwyptiaregvvdhrhgevtvdgipvl 204
Qy 269 --SCEELGNMVOELSLGLVNLVNOQPSENLKRVSNDOQFELWELIGPPKTRNMSACWQDGRF 326
Db 205 ngdceda--iarslsdlalvklirevahrqaeiyl-----rm 242
Qy 327 FAENETWVVDSCITCT-CKFKFTICHQITC-PPATCASPSPFVEGECC-----PSCLHSVD 379
Db 243 llen-----cagcknplttndnglrliepdcrsanpcygvvecidsaagprcgh--- 289
Qy 380 GEEGWSFWAEWTQCSVTGSGSTQQRGRSCDVTNTCLGSPISQTRACSLSKDTRIRODGG 439
Db 290 -----cpigfigdksck-----pgv---tcahnmcypgvq--- 317
Qy 440 WSHNSPWSGCVTGVGNITRILCNRPVPMQMGKCKGSGRETAKCQAGPCPIDGRWSP 499
Db 318 -----chdtv-----ngaqc----- 328
Qy 500 WSPWSACTVTCAGGIRTRVCNSPEPOYGGKACVGDVQERQCMKRCSCPVDGCLSNPCF 559
Db 329 -----scpagyegdgrtcslnp-----cldtcp 353
Qy 560 PGAQC--SSPFGSGWSGSCPGFLNGTHCEDLDECALVPDICFSTSKVPRCVNTOPGF 617
Db 354 sgaclqvgyp-yfhciscpmghevngtscrdnnec-illypc---delatctnlspgf 408
Qy 618 HCLPCPPRYRGNPQVGVGLE--AAKTEKQVCEPENPCDKTHNCKHAECT----- 666
Db 409 qcpcpvgfdgthahgyfadyysvyrqtclvdcertgfrfrcpehstcineigsrycq 468
Qy 667 -YLGHFSDPMY-----KCEQTGYAGDGLIGED 694
Db 469 chegyvtngtycldrssvfmcpgdgtvcdnavclrmnirhkchcnvgwagnglicgrd 528
Qy 695 SLDLGNLNLVLCATNATYHCIKDKONCPHLNPSGOEDFKDGIGACDDEDDNDGCVTEKD 754
Db 529 tdvsgwpaqagc---pelrcqrdrncpklnpsggadadlghgvgdcdadgdnvqnsqd 585
Qy 755 NCQLLFNPRQADYDKDEVGRDCNCPYVHNPAQITDNNEGDACSVDIDGDDVFNERN 814
Db 586 ncwlaynteqlsdgdkvgdvcnclvknprqltdedglgdecgdldndsipnaldn 645
Qy 815 CPYYVNTDQRTDGDGVGDHCDNCPVHNPDQTDVNDLVGDQCDNNEIDDDGHQNNQD 874
Db 646 cpllpnpsqsdvndgvgdadcncpnlpnqdkrdmndfvgdchrldldgddgvpnsld 705
Qy 875 MCPYISNANQADHDDRGDGDACDDDDNDGVDPDDRCRLVFNPDQEDLDGDRGDIKD 934
Db 706 ncpmvsnsdqltdgdtgdecddmddgdgipnykdnopkpnkqddfrnngkdsced 765
Qy 935 DFNDNDIPDIDVCPENNAISETDFRNFMVPLDPKGTQIDPNWVIRHOGKELVQFANS 994
Db 766 dedvvgvngmdncpnsmihhtdfrltlqtlpdkglsqadpnwvvhangteivqlns 825
Qy 995 DPGTAGVDFEFGSVDFSTFYVNTDRDDVAGFFVFGQSSSRFYVVMKWQVQTYWEDQP 1054
Db 826 dpgiavgkdaifggvdfgtfyndtdtdddvagfvsygsykyvvyvkkgtqtywep 885
Qy 1055 TRAVYGSVSLKVVNNTGTGHEHLNALWHTTGGVVRTLWHPDRNWKDYTAIRWHL 1114
Db 886 ftsaaepgiqlklvnstecgppmnmnswhnegntdgarllwkdplnawkertsyws 945
Qy 1115 THRPKTYIRVLVHEGQVMADSGPIYDOTYAGAGRLGLFVFSQBMYFSDLYKECDR 1171
Db 946 vhrpaiglrilqmhagnrlifdsgnvfdstlkgrlgvfcfsqrmlwslnlykenn 1002

RESULT 14
AAB58933
ID AAB58933 standard; Protein; 444 AA.
XX AC AAB58933;

CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 444 AA;

| | | | | | |
|-----------------------|------|---|--------------------|-----------|-------------|
| Query Match | | 21.6% | Score 1428; | DB 22; | Length 444; |
| Best Local Similarity | | .60.7% | Fred. NO. 2.4e-80; | | |
| Matches 258; | | Conservative 47; | Mismatches 116; | Indels 4; | Gaps 1; |
| QY | 751 | DEKDNQCLLFNPRQADYDKDEYDCDCNCPYVHNPAQIDTDNNGEGDACSVDDIDGDDVFN | 810 | | |
| Db | 1 | neqdnocvlihdvdqrnsdkdlfgdacdcnclsvlxndqktdtdggrgdacddmdgdgikn | 60 | | |
| QY | 811 | ERDNCPPYVNTDQRTDGVGDHCDNCPLVHNPDQTDVDNDLVGDCDNNEDIDDDGHQ | 870 | | |
| Db | 61 | lildncpkfnpqrqdkdggvgdacsdpvsnpnqsdvndnlvgdsctnqdsdgdghq | 120 | | |
| QY | 871 | NNQDNCPIYSNANQADHDGDCGDCADPDDDNDGVDP---- | 926 | | |
| Db | 121 | dstancptvinsaqidtdkdgiqdecdaddndgipdivppgpdcncrlvpnpagedsnsd | 180 | | |
| QY | 927 | GRGDIKDDFDNDTIPDIDVCPENNAISETDFRNFOMVPLDPKGTTOIDPNWVIRHOGK | 986 | | |
| Db | 181 | gvgdicesdfddqvidridvcpenaevlttdfrayqvtvldpegdaqidpnwvvingm | 240 | | |
| QY | 987 | ELVQTANSOPGIAVGDFDEGSVDFSGTFYVNTDRDDYAGVFYGSSSRFYVVMKQVT | 1046 | | |
| Db | 241 | eivqtmsdpglavgytafngydfegtfnvtqtdddyagfigygdsssfyvvmmwkate | 300 | | |
| QY | 1047 | QYWEQPTFRAYGSGVSLKVNSTGTCEHLRNALWHTGQVYRTLWHDPRNIGWKD | 1106 | | |
| Db | 301 | qtywqatpfravaepgiglkavkstktpgehlrnsilwhtgdsdqvrllwkdsrnvgwk | 360 | | |
| QY | 1107 | YTAIRWHLTHRPKTYIRVLVHEGQVMADSGPIYDQTYAGGRLGLFVFSQEMVYFSDLK | 1166 | | |
| Db | 361 | kvsyrfwflqhrpqgyirvrfyegselvadsgvtidttmrggrlgvfcfsqeniwsnlk | 420 | | |
| QY | 1167 | YECRD | 1171 | | |
| Db | 421 | yrncnd | 425 | | |

Search completed: August 9, 2002, 10:01:16
Job time: 262 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 09:58:54 ; Search time 17.01 Seconds
(without alignments)
1682.939 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWVWPSTQAGH.....RVFSQEMVYFDLKYECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 6570 | 99.5 | 1172 | 1 | US-08-313-288B-19 |
| 2 | 4228 | 64.0 | 1170 | 1 | US-08-313-288B-20 |
| 3 | 1899 | 28.8 | 889 | 5 | PCT-US93-11725-2 |
| 4 | 1890 | 28.6 | 961 | 5 | PCT-US93-11725-4 |
| 5 | 880 | 13.3 | 239 | 5 | PCT-US93-01652-1 |
| 6 | 783 | 11.9 | 441 | 3 | US-08-985-526-3 |
| 7 | 713 | 10.8 | 218 | 3 | US-08-985-526-1 |
| 8 | 415 | 6.3 | 469 | 1 | US-08-313-288B-15 |
| 9 | 399 | 6.0 | 788 | 2 | US-08-918-914-4 |
| 10 | 388.5 | 5.9 | 810 | 2 | US-08-820-170A-34 |
| 11 | 388.5 | 5.9 | 810 | 3 | US-09-055-699-34 |
| 12 | 388.5 | 5.9 | 810 | 4 | US-09-273-565-34 |
| 13 | 388.5 | 5.9 | 810 | 4 | US-09-565-538-34 |
| 14 | 361.5 | 5.5 | 816 | 2 | US-08-820-170A-37 |
| 15 | 361.5 | 5.5 | 816 | 3 | US-09-055-699-37 |
| 16 | 361.5 | 5.5 | 816 | 4 | US-09-273-565-37 |
| 17 | 361.5 | 5.5 | 816 | 4 | US-09-565-538-37 |
| 18 | 329.5 | 5.0 | 807 | 1 | US-07-862-021B-10 |
| 19 | 329.5 | 5.0 | 807 | 1 | US-08-313-288B-10 |
| 20 | 329.5 | 5.0 | 807 | 5 | PCT-US93-03164-10 |
| 21 | 326.5 | 4.9 | 2523 | 1 | US-08-185-432-18 |
| 22 | 322 | 4.9 | 802 | 1 | US-07-862-021B-12 |
| 23 | 322 | 4.9 | 802 | 1 | US-08-313-288B-12 |
| 24 | 322 | 4.9 | 802 | 5 | PCT-US93-03164-12 |
| 25 | 314.5 | 4.8 | 2556 | 1 | US-08-185-432-17 |
| 26 | 312 | 4.7 | 2471 | 1 | US-08-185-432-16 |
| 27 | 312 | 4.7 | 2471 | 1 | US-08-083-590A-19 |

28 312 4.7 2471 3 US-08-532-384-19 Sequence 19, Appl
29 306.5 4.6 2556 1 US-08-083-590A-20 Sequence 20, Appl
30 306.5 4.6 2556 3 US-08-532-384-20 Sequence 20, Appl
31 302.5 4.6 2703 1 US-08-185-432-19 Sequence 19, Appl
32 296 4.5 568 1 US-07-862-021B-14 Sequence 14, Appl
33 296 4.5 568 5 PCT-US93-03164-14 Sequence 14, Appl
34 290 4.4 943 2 US-08-808-982-7 Sequence 7, Appl
35 290 4.4 943 4 US-09-306-902A-7 Sequence 7, Appl
36 286.5 4.3 898 2 US-08-808-982-5 Sequence 5, Appl
37 286.5 4.3 898 4 US-09-306-902A-5 Sequence 5, Appl
38 278.5 4.2 717 3 US-08-872-855-9 Sequence 9, Appl
39 276 4.2 1219 4 US-08-882-046-5 Sequence 5, Appl
40 272 4.1 728 4 US-08-981-392-2 Sequence 2, Appl
41 270 4.1 729 3 US-08-872-855-8 Sequence 8, Appl
42 268 4.1 60 1 US-07-646-531D-12 Sequence 12, Appl
43 268 4.1 60 2 US-08-488-273-12 Sequence 12, Appl
44 268 4.1 60 6 5426100-12 Patent No. 5426100
45 267 4.0 1193 2 US-08-400-159-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40038-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-313-288B-19

Query Match 99.5%; Score 6570; DB 1; Length 1172;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MWRLVLLALWVWPSTQAGHODKDTTDLFSLISINRKTIGAKQFRGPDGVPAYRFRVF 60
Db 1 MWRLVLLALWVWPSTQAGHODKDTTDLFSLISINRKTIGAKQFRGPDGVPAYRFRVF 60

```

QY 61 DYIPPNADLSKTIKIMROKGGFFLTAQLKODGKSRGTLTALLEGPSLQSOFEIVSNP 120
DB 61 DYIPPNADLSKTIKIMROKGGFFLTAQLKODGKSRGTLTALLEGPSLQSOFEIVSNP 120
QY 121 ADTLDLTWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCGLDLSFALDEFFY 180
DB 121 ADTLDLTWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCGLDLPVALDEFFY 180
QY 181 EHLQAEKSMYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQOAGAEINSEN 240
DB 181 EHLQAEKSMYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQOAGAEINSEN 240
QY 241 TETLRGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARVND 300
DB 241 TETLRGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARVND 300
QY 301 NOFLMELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTICHQITCPPATC 360
DB 301 NOFLMELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTICHQITCPPATC 360
QY 361 ASPSFVEGECPCSLHSDVDEGEGSWPABWTCQSVTCGSGTOQRGSCDVTSTNTCLGPSI 420
DB 361 ASPSFVEGECPCSLHSDVDEGEGSWPABWTCQSVTCGSGTOQRGSCDVTSTNTCLGPSI 420
QY 421 QTRACSLSKCDRIQDGGWSHSPSSCSVTGCGVGNITRIRLCHNSPVPQMGKCKGSG 480
DB 421 QTRACSLSKCDRIQDGGWSHSPSSCSVTGCGVGNITRIRLCHNSPVPQMGKCKGSG 480
QY 481 RETKACQAPCIDGRWSPSPSACTVTCAGGIRERTRVCNSPFPQYGGKACVGDVQER 540
DB 481 RETKACQAPCIDGRWSPSPSACTVTCAGGIRERTRVCNSPFPQYGGKACVGDVQER 540
QY 541 QMCNRSCPDCLNCPFGAGCSPFPGSWSCSCSPVFGNGTHCEDLDECALVPDI 600
DB 541 QMCNRSCPDCLNCPFGAGCSPFPGSWSCSCSPVFGNGTHCEDLDECALVPDI 600
QY 601 CFSTSKVPRCVNTQGFHCLPCPPRYRGQNPVGVGLEAAKTEKQVCEPENCKDKTHNCH 660
DB 601 CFSTSKVPRCVNTQGFHCLPCPPRYRGQNPVGVGLEAAKTEKQVCEPENCKDKTHNCH 660
QY 661 KHAECIYLGHSFDPMYKCEQTYAGDGLICGEDSDLDGWPNLVGCATNATHYHCKNC 720
DB 661 KHAECIYLGHSFDPMYKCEQTYAGDGLICGEDSDLDGWPNLVGCATNATHYHCKNC 720
QY 721 PHLPNSGQEDFKDGIADCDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780
DB 721 PHLPNSGQEDFKDGIADCDDDNDGVTDEKNCOLLFNPRQADYDKDEVGDRCDNCP 780
QY 781 YVHNPAQIDTDNNGEGDASVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHDCNCP 840
DB 781 YVHNPAQIDTDNNGEGDASVDIDGDDVFNERNDCPYVYNTDQRTDGDGVGDHDCNCP 840
QY 841 VHNPDQTDVNDLVGQDCNNEDIDDDGHQNDQNCPIYISNANQADHDHDEGGDADCDPD 900
DB 841 VHNPDQTDVNDLVGQDCNNEDIDDDGHQNDQNCPIYISNANQADHDHDEGGDADCDPD 900
QY 901 DNDGVPDDRNCRLVFNPDQEDLDGDRGDCDDFDNDNIPIIDVCPENNAISETDPR 960
DB 901 DNDGVPDDRNCRLVFNPDQEDLDGDRGDCDDFDNDNIPIIDVCPENNAISETDPR 960
QY 961 NFQMPVLPKGTQIDPNVNIHQKELVQTANSDPGIAVGFDEFGSVDFSGTFYVNTDR 1020
DB 961 NFQMPVLPKGTQIDPNVNIHQKELVQTANSDPGIAVGFDEFGSVDFSGTFYVNTDR 1020
QY 1021 DDDYAGFVFGYSSSRFYVMKQYTOTYVWEDQTPRAYGCVSLKVNSTTGTGEHLRN 1080
DB 1021 DDDYAGFVFGYSSSRFYVMKQYTOTYVWEDQTPRAYGCVSLKVNSTTGTGEHLRN 1080
QY 1081 ALWHYGTNTPGQVRLWHDPRNIGWKDYATYRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
DB 1081 ALWHYGTNTPGQVRLWHDPRNIGWKDYATYRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140

```

```

QY 1141 YDOTYAGRGRLGLFVSQEMVYFSDLKYECRDI 1172
DB 1141 YDOTYAGRGRLGLFVSQEMVYFSDLKYECRDI 1172
RESULT 2
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313.288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-288B-20
Query Match 64.0%; Score 4228; DB 1; Length 1170;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 725; Conservative 170; Mismatches' 267; Indels 14; Gaps 6;
QY 1 MYWRL-VLLALWVPSTQAGHODKD-TTFDLFSINIRKKTIGAKQFQPDGVPAYRRV 58
DB 3 LAWGGVFLMHVCTNRIPESGGNSVDFIFELGAARKSGRRLYKGPDPSSAFRIE 62
QY 59 RFDYIPPNADLSKTIKIMROKGGFFLTAQLKODGKSRGTLTALLEGPSLQSOFEIVSN 118
DB 63 DANLIPPVDDKFDQDLVDAVTEKGFLLASLURQKKTGTLLALERKDHSGQVFSVSN 122
QY 119 GPADTLDLTWIDGTRHVVSLDVLGADSQWKNVTVQVAGETYSLHVGCGLDLSFALDEP 178
DB 123 KAGATLDLSLTAVQGHVSVVEALLATGQWKSITLQVQEDRAQYIDCEKEMNAELDVP 182
QY 179 F---YEHQAERSMYAKSARESHFRGLLQNVHLVFNSEVEDILSKKCGQOQOAGAEIN 235
DB 193 IQSVTRDLASTARLAKGGVND-NFQGLQNVRFVFGTTPEDILRNKSGSSSTVLLT 241
QY 236 AISENTETLRLGPHVTTEYVGPSSERPEVCERSCBELGNMVOELSLHLVNPQSENARV 295
DB 242 L--DNNVYNGSSPAIRTNVIGHKTKDLQAICIGISCDLSSMWLELRGLRTIVTTLQDSIR 299
QY 296 RVSNDNQFLWELIGGPPKTRNMSACWQDGRFFAENETWYVDSCTTCTCKFKTICHQITC 355

```



```
Db 300 KYTEENKELANERLRPP-----LCHNGVQYRNNEWVDSCTECHCONSVTICKKVSVC 353
QY 356 PPATCASPFEVEGCCPSCLSHSYDGBEGSWPAEWTQCSVTGSGTQQRGRSDVTSNRC 415
Db 354 PIMPCSNATVPDECCPCRCWPSDSADGNSPWESEMTSCSTSCGNGIQQRGRSDSLNNRC 413
QY 416 LGPSIOTRACLSKCDTRITRODGGWHSWSPSSCSTCGVGNITRILCLNSPVPQMGKN 475
Db 414 EGSSVOTRCHIQEOKRQKQDGGWHSWSPSSCSTCGDGVITRILCLNSPSPQMGKP 473
QY 476 CKSGRETRACQACAPCIDGRNWPSPWASACTVTCAGGIRERPRVCSPEPQYGGKACVG 535
Db 474 CESEARETRACKDACPINGGWPSPWDICSVTCGGVQKRSRLCNNPTPQFGGKDCVG 533
QY 536 DVQEROMKNRSCVPDGLCLSNPCFPAGQSSFPDGSWSCGSCPVGFLNGTHCEDLDECA 595
Db 534 DVTENQICNKQDCPIDGLCLSNPCFAGVKCTSYPDGSKWKGACPPGYSGNGIQCTDVECK 593
QY 596 LVPDICFSTSKVPRCVNTOPGFHCLPCPRYRCNQPVGVGLEAAKTEKOVCEPENCKDK 655
Db 594 EVDPAFCNHNHGERCENTDPGYNCLCPFRFTGSGPQGVQGHATANKQVCKRPNCTDG 653
QY 656 THNCHKHAECIYLGHFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCATNATYHC 715
Db 654 THDCNNAKCNLYLGHYSDPMYRCECKPGYAGNGIICGEDTDLGWNENLYCVANATYHG 713
QY 716 IKDNCPLHNSGQDEDFKQIGDACDDDDDDNGVTDKQNCQLLFPNROADYDKDEVDGR 775
Db 714 KKDNCPLNPSGQEDYDKQIGDACDDDDDDNDKIPDRDNCPPHYNPAQYDYDRDDVGR 773
QY 776 CDNCPIVHNPAQIDTNNEGDACSVDIDGDDVFNERNDCNCPYYVNTDQRTDGDGVDHC 835
Db 774 CDNCPIVHNPAQIDTNNEGDACAADIDGDLNERDNCQYVYVNDQRTDMDGVDQC 833
QY 836 DNCPLVHNPDQTDVNDLVGDCDDNEDIDDGHNQNDNCPIYISANQADHRRDQGDGA 895
Db 834 DNCPLHNPDQDSDSDRIGDTCNNQNDIDEGHQNLLDNCPIVYVNPANQADHDKGDA 893
QY 896 CDPDNDGVDPRDRDNCRLVFPDQEDLDGDRGDKDCKDDFDNDNIPDIDVCPENNATS 955
Db 894 CDHDDNDGTPDDKNCRLVFPDQKDSGDRGDAKDDFDHDSVPDIDVCPENNATS 953
QY 956 ETDFRFQMPVLPKGTQTDPMWVLRHOGKELVOTANSDPGTAVGDFEGSFGSTFY 1015
Db 954 ETDFRFQMPVLPKGTQNDPMWVLRHOGKELVQVNCDPGLAVGYDEFNAVDFSGTF 1013
QY 1016 VNTDRDDYAGFVGYQSSSRFYVVMKQVTOYFWDQTPRAYGSGVSLKVVNSTGTG 1075
Db 1014 INTERDDYAGFVGYQSSSRFYVVMKQVTOYFWDQTPRAYGSGVSLKVVNSTGTG 1073
QY 1076 EHLRNALWHTGNTPGQVRLTLDHDPNRIGWKDYATYRWHLTTHRPKTYIRVLVHEGKVMA 1135
Db 1074 EHLRNALWHTGNTPGQVRLTLDHDPNRIGWKDYATYRWHLTTHRPKTYIRVVMYEGKINA 1133
QY 1136 DSGPIYDQYAGRGLGLFVSQEMVYFSDLYKECRD 1171
Db 1134 DSGPIYDKTYAGRGLGLFVSQEMVYFSDLYKECRD 1169
```

RESULT 3

```
PCr-US93-11725-2
; Sequence 2, Application PC/TUS9311725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMAN THROMBOSPONDIN-4
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
```

```
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch
; COMPUTER: IBM-compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11725
; FILING DATE: filed herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/985,296
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/77005WO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 889 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: yes
; PCT-US93-11725-2
```

Query Match 28.8%; Score 1899; DB 5; Length 889;

Best Local Similarity 51.9%; Pred. No. 4.5e-139;

Matches 348; Conservative 79; Mismatches 184; Indels 60; Gaps 7;

```
QY 549 PVDGCLSNPCFPAGQSSFPDGSWSCGSCPVGFLNGTHCEDLDECALVDPICFSTSKVP 608
Db 215 PRPCRDATSCFRVRCID-TEGFGQCPPEGYTGNGVICTDYDECLNP--CFLG---V 268
QY 609 RCVNTQPGFHLCPPPRYRGNGPVGVGLEAAKTEKQV----- 645
Db 269 RCINTSPGFKCESPPGYTGSTIQIGINFQKQNKQVCTDNECENGRNGGCTSNSLCIN 328
QY 646 -----CEPENCKDKTHNCHKHAECIYLGHFSDPMYKCECOTGY 684
Db 329 TMGSFRGCKPGYVGDQIKGCKPEKSRHGNPNCHASAQC-----SEBKGDVTCVSGW 385
QY 685 AGDLGICGEDSLDGLWPNLNLVCATNATYHCIKDNCPLHNSGQEDFDKDGIGDADDD 744
Db 386 AGNGYLCGRDITDIDGYPDEALPCPK---NCKDNCVYVNPNSGQEDTDKDNIGDCEDA 442
QY 745 DNDGVYDEKDNCLLPNROADYDKDEVGDRDNCPIVYVNPANQADHRRDQGDGA 804
Db 443 DGDGILNEQDNCVLAANIDQKNSDQDIFGDACDNCRLTLNNDQRTDNDGKGDACDDMD 502
QY 805 GDDVFNERNDCPIVYVNTDQRTDGDGVDHCDNCPLVHNPDQTDVNDLVGDCDDNEDI 864
Db 503 GDGIKNILDNCQVNPVNDQKDGVDGDIICDSCPDIINPNQSDINDLVGDCDDTNDQS 562
QY 865 DDDGHONNDNCPIYISANQADHRRDQGDGACDPDDNDGVPD---DRDNCRLVFNPDQ 920
Db 563 DGDGHQDSTDNCPTVINSQOLDKDGIGDECDDDDDNDGIPDTPVPPDNCRLVNPQG 622
QY 921 EDLDGDRGDKICKDDFDNDNIPDIDVCPENNAISETDFERNFQVPLDPKGTQIDPNWV 980
Db 623 EDDNDGVDGVDCEADFDQDTVIDRIVDVCPENAEITLTDPRAYQTVVLDPGDAQIDPNWI 682
QY 981 IRHQKELVOTANSDPGIAVGDFEFGVSDFSGTFYVNTDRDDYAGFVGYQSSSRFYV 1040
Db 683 VLNQGMIEVQTMNSDPGLAVGYTAFNGVDFEGTFHVTMTDDDYAGFIFGYQSSSFYV 742
QY 1041 MKQVTOYFWDQTPRAYGSGVSLKVVNSTGTGHEHLRNALWHTGTGQVFTLHDPNR 1100
Db 743 MKQTEQTTWQATPFRFAVEPGLQKAVKSKSGFGEHLRNALWHTGTDNDQVRLWKDPR 802
QY 1101 NIGWKDYATYRWHLTTHRPKTYIRVLVHEGKVMAADSGPIYDQYAGRGLGVFSQEMV 1160
Db 803 NVGKDKVSRWFLQHRPQVGYIRARFYETGELVADSGVTVDTRMRGRLGVFCFSQENI 862
```

Qy 1161 YFSDLKYCRD 1171
Db 863 IWSNLKYRCND 873

RESULT 4
PCT-US93-11725-4
; Sequence 4, Application PC/TUS9311725
; GENERAL INFORMATION:
; APPLICANT: HUMAN THROMBOSPONDIN-4
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch
; COMPUTER: IBM-compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11725
; FILING DATE: filed herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/985,296
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7005W0
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 961 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: yes
PCT-US93-11725-4

Query Match 28.6% Score 1890; DB 5; Length 961;
Best Local Similarity 53.1%; Pred. No. 2.5e-138;
Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps 9;

Qy 549 PVDGCLNPFPGACQSFPGNSGSCPVGFLNGTHCEDLDLCAVPDIFCSTSKVP 608
Db 286 PPRCDNSNPCFRGVQCTSDRDC-FQCGPCPGYTGNGITCIDVDECKYHP--CYPG---V 339

Qy 609 RCVNTPQGFHCLPCPRYRGNQPVGVGLEAAKTEKQVC-----EPEN----- 650
Db 340 HCINLSPGFRDACPVGFTGPMVQGVGIFSAKSNKQVCTDDECRNGACVNSICVNTLG 399

Qy 651 -----PK-----DKTHNCHHABCIYLGHPFSDPMYKCEQCTGYAG 686
Db 400 SYRCGPCPGYTGDIQCKKVERCNPNELNCPNSVNAQCI---EERQGDVTCVGVGAG 456

Qy 687 DGLIGESDLDGWNPLNLVCAATNATYHCIDNCPHLPNSGQEDPDKGIGDADDDDDN 746
Db 457 DGYICGKDVIDSYPDEELPCASAR---NCKKDNCKYVNSGQEDADRDGIGDADDEDADG 513

Qy 747 DGVTEKDNKCOLLFPNQADYDKDEVGRDCNCPVHNAQITDNDNGEGDACSVDIDGD 806
Db 514 DGLNEQDNCVLHNVQDNRNSDKDIFGDACDNCNLSVLNNDQKDTDGDGKGDACDDMDGD 573

Qy 807 DVNERDNCPPVYNTDQRDTDGDGVDGDCNCPVHNPQDQDNDLVGDGDCDNDNEDIDD 866
Db 574 GIKNILDNCPFPNRQDKDGDGVDGADCDSCPDVSNPNQSDVDNDLVGSDCDTNQSDG 633

Qy 867 DGHONNDCNPVSNANOADHROGOGDADCPDDDDNDGVPD-----DRNCRLVENPDDOED 922
Db 634 DGHQDSTNCPTVINSALQDTDKDGIGDECDDDDDNDGIPDLVPPGPNCRLYPNPAQED 693

Qy 923 LDGGRGDICKDDFDDNDNIPDIDDVCPENNAISETDFRNFQVPLDPRKGTQIDPNWVIR 982
Db 694 SNSDGVGDICESDFDQDQVIDRIDRVCENAEVTLTDFRAYQTVGLDPEGDAQIDPNWVVL 753

Qy 983 HQKRELVTQANSDFCIAGVDFEFGSVDFSVFVNTDRDDYAGVFCYQSSSREYVVMW 1042
Db 754 NOGMEIVQTMNSDPLAVGYTAFNGVDGEGTFHVNTQDDDYAGFIFGYQSSSSEYVVMW 813

Qy 1043 KQVQTQYWEQDPTRAYGYSVSLKVNSTTGTGHEHLRNLWHTGTQGVRTLWHDPRNI 1102
Db 814 KQTEQTYWQATPFRAVAEPGQLKAVKSKTGPGEHLRNLWHTGTDSDQVRLWKDSRV 873

Qy 1103 GWKDYATYRWHLTHRPKTYIRVLVHEGKQVWADSGPIYDQTYAGRGRLGLFVFSQEMVYF 1162
Db 874 GWKDKVSYRWFLQHRPQVGYIRVREYEGSELVADSGVITDTTMRGRLGVFCFSQENIIV 933

Qy 1163 SCLKYECRD 1171
Db 934 SNLYRCND 942

RESULT 5
PCT-US93-01652-1
; Sequence 1, Application PC/TUS9301652
; GENERAL INFORMATION:
; APPLICANT: Bouck, Noel P.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Good, Deborah J.
; APPLICANT: Frazier, William A.
; TITLE OF INVENTION: Method and Composition for
; TITLE OF INVENTION: Inhibiting Angiogenesis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
; STREET: 100 South Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01652
; FILING DATE: 19930222
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,656
; FILING DATE: 24-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/464,369
; FILING DATE: 12-JAN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fentress, Susan B.
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92005-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)-456-8000
; TELEFAX: (312)-456-7776
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
PCT-US93-01652-1

Query Match 13.3%; Score 880; DB 5; Length 239;
Best Local Similarity 58.4%; Pred. No. 9.9e-61;
Matches 139; Conservative 35; Mismatches 64; Indels 0; Gaps 0;

QY 315 RNMSACWQDGRFAENETWVDSCTTCTCKKFKTICHQITCPPATCASPFEVEGCCPSC 374
DB 2 RRPPLCYHNGVQYRNNEWTVDSCTECHQNSVTICKKVCSCIPMPCSNATVPDGECCPRC 61
QY 375 LHSVDEEGSWPAEWTCQSVTCGSGTQQRGRSCDVTSNTCLGPSIQTRACSLSKCDTRI 434
DB 62 WPSDSADGWSPEWTSCTSCGNGIQQRGRSCDLSNNRCEGSSVQTRTCHIQCCKRF 121
QY 435 RQDGGHSHSPSSCSVTGCGVGNIRILCNPSVPOMGKNGKSGRETKACQAGPCPID 494
DB 122 KQDGGHSHSPSSCSVTGCGVGNIRILCNPSVPOMGKNGKSGRETKACQAGPCPID 494
QY 495 GRWSPWSPSACTVTCAGGIRERTRVCNPSPEQYGGKACVGVQVROMENKRCSPVDG 552
DB 182 GGWGPSPWDICSVTCGGGVQKRSRLCNPAFQFGLDCVGDVTENQICNKQDCPIDG 239

RESULT 6
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/985,526
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-3

Query Match 11.9%; Score 783; DB 3; Length 441;
Best Local Similarity 39.9%; Pred. No. 8.1e-53;
Matches 159; Conservative 57; Mismatches 134; Indels 48; Gaps 12;

QY 138 SLEDVGLADSQKNVTVQVAGETYSLHVGCDLIDFALDEPPEY---HLQAEKSRMYAK 194
DB 77 SADDGSPWSEWTSCTSCGNGIQQRGRSCDLSNNRCEGSSVQTRTCHIQCCKRF---K 133

QY 195 GSARESHFR-----GLLQNVHLVFENSVEDILSKGCGQCGAGAEINAISENTET 243
DB 134 QDGGHSHSPSSCSVTGCGVGNIRILCNPSVPOMGKNGKSGRETKACQAGPCPID 190
QY 244 LRLGPHVTTEYVGPSSERRP-EVCSRSCEBGLNMVQELSLGLVHLVNPQSENKRVSNQ 302
DB 191 INGG-----WGPWS---PMDICSVTC---GGGVQKRSRLCVDSTRMTEEN-KELANE-- 234
QY 303 FLWELIGGPCKTRNMSACHQDGRFAENETWVDSCTTCTCKKFKTICHQITCPPATCAS 362
DB 235 -----LRRPP-----LCYHNGVQYRNNEWTVDSCTECHQNSVTICKKVCSCIPMPCSN 283
QY 363 PSFVEGCCPCLSHSVDEEGSWPAEWTCQSVTCGSGTQQRGRSCDVTSNTCLGPSIQTR 422
DB 284 ATPDGECCPRWSDSADGWSPEWTSCTSCGNGIQQRGRSCDLSNNRCEGSSVQTR 343
QY 423 RACSLSKCDTRIQRDGGHSHSPSSCSVTGCGVGNIRILCNPSVPOMGKNGKSGRE 482
DB 344 RTCHIQCCKRFKQDGGHSHSPSSCSVTGCGVGNIRILCNPSVPOMGKNGKSGRE 403
QY 483 RTACOGAPCPIDGRWSPWSPSACTVTCAGGIRERTRV 520
DB 404 TRACKKDACPINGWGPSPWDICSVTCGGGVQKRSRL 441

RESULT 7
US-08-985-526-1
; Sequence 1, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/985,526
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-1

Query Match 10.8%; Score 713; DB 3; Length 218;
Best Local Similarity 52.2%; Pred. No. 8.2e-48;
Matches 117; Conservative 35; Mismatches 56; Indels 6; Gaps 1;

QY 297 VSNDRNQFLWELIGGPCKTRNMSACHQDGRFAENETWVDSCTTCTCKKFKTICHQITCP 356
DB 297 VSNDRNQFLWELIGGPCKTRNMSACHQDGRFAENETWVDSCTTCTCKKFKTICHQITCP 356

Db 1 MTEENKELANELRRPP-----LCYHNGVQYRNNEETVDSCTECHCONSVTICKKVSCLP 54
QY 357 PATCASPVEGECPCSLHSDVDEGWSWPAEWTCQSVTCGSGTQQRGRSCDVTNTCL 416
Db 55 IMPCSNATVPDECCPCRPWSDSADGHSWSEWTSCTSCGNGIQQRGRSCDLSNRCE 114
QY 417 GPSIOTRACSLKCDTRIQDGGSHWSPWSSCVTCGVGNITRIRLNSVPVQMGKNC 476
Db 115 GSSVOTRCHIOECDKRFKQDGGSHWSPWSSCVTCGDGVITRITNLCSPOMNKPC 174
QY 477 KSGRETKACQACPIDGRSPWSPWACTVTCAGGIRERTR 520
Db 175 EGREAETACKRADCPGNGWGPWSPWDICSVTCGGGVOKRSRL 218

RESULT 8
US-08-313-288B-15
; Sequence 15, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avi Hu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
; LENGTH: 469 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-313-288B-15

Query Match 6.38; Score 415; DB 1; Length 469;
Best Local Similarity 25.18; Pred. No. 3.4e-24;
Matches 124; Conservative 32; Mismatches 150; Indels 188; Gaps 20;

QY 360 CASPSFV----EGECPCSLHSDVDEGWSWPAEWTCQSVTCGSGTQQRGRSCDVTNTCL 415
Db 57 CLNTAFATQKRSGLQCPC-----RSPRSLWSTWAPCVTCSEGSQRLRYRCVGNWQC 111
QY 416 LGP-----SIQTRACSLKCDTRIQDGGSHWSPWSSCVTCGVGNITRIRLNSVPV 469
Db 112 SGKVAPGTLEWLQACEQOCPEM---GGWSGWGPWPCSVTCSGKTRRRACNHP 168
QY 470 QMGKNCKGSGRETRAC-----QGA----- 489

Db 169 KCGG-HCPGQAQSEBACDQOVCPTHGAWATGWPTPCASCHGHPHPKETRSRKCSAP 227
QY 490 -----PCPIDGRWSPWSPWACTVTCAGIRTRVCNSP 524
Db 228 BPSKPKPKPCGLAYEQRRTGLPPCPVAGWGPWGPVSPCPVTCGLGQTMQRTCNHP 287
QY 525 EPQYGGKACVGDVQERQCNKR-SCPVDGCLS-----NPC-----FPGAQCSFDPDGSWSC 574
Db 288 VPQHGPPCAGDATRTHICNTAVPCVPDGDWDSWGEWSPCIRRNKMSISCQEIP-GQOSR 346
QY 575 GSCPVGFLNG-----THCEDLDECALVPD-----ICFSTSKYPRCVNTOPG 616
Db 347 GTRCRGPKFGRHRCAGQQODIRHCYSIQHCLPKGSWSWSTWGLC-----MPPCGPNPTR 401
QY 617 FH---CLPCPPRYRGNQPVGVGLEAAKTEKQVCEPENPCDKDTHNCHKHAECIYLGHFSD 673
Db 402 ARQLCTPLLPKY----PPTVSMVGEQGEKNVT-----FWGR 434
QY 674 PMYKCECOTGYAGDGLICGEDSDLDGWPNLNLVCAATNATYHCIKDNCPLPNSGOEDFDK 733
Db 435 PLPRCE---ELQGGKLVVEE-----KRPLHVP----- 459
QY 734 DGIGDACDDDDND 747
Db 460 -----ACKDPBEE 468

RESULT 9
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

QY 832 -----GDHCNCPVHNPDQTDVNDLVGDCDNE---DIDDDGHQNN 872
Db 779 SVWTMAGSPCTTC-----KCKNGRVCCSVDFECLQNN 810

RESULT 11

US-09-055-699-34

; Sequence 34, Application US/09055699

; Patent No. 6005088

; GENERAL INFORMATION:

; APPLICANT: Tsutomu, FUJIWARA

; APPLICANT: Takeshi, WATANABE

; APPLICANT: Masato, HORIE

; APPLICANT: Toyomasa, KATAGIRI

; TITLE OF INVENTION: HUMAN GENE

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: United States

; ZIP: 20037-3202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/055,699

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/820,170

; FILING DATE:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 293-7060

; TELEFAX: (202) 293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 810 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-055-699-34

Query Match 5.9%; Score 388.5; DB 3; Length 810;
Best Local Similarity 20.6%; Pred. No. 8.3e-22;
Matches 208; Conservative 106; Mismatches 347; Indels 349; Gaps 50;

QY 8 LALVWV-----PSTQAGH-QDKDTTFDLFSISNINRKTIGAKQFQPGDPGVPAYRFVREF 61

Db 5 LILVWFVCFVARTVWGMDPDQLQMDIVTDLVNTLGLVAQVSGMHNASKAFLODIE 64

QY 62 ----YIPVNDLSKITKMKQEGFLLTAQLKDGKSRGTLALLEGPLSQRFELVS 117

Db 65 REIHAAPIVS-----EKLIQLPQNKSEITLATVQOKPSTSGVILSIR--ELEHSYFELES 118

QY 118 NGPADTLDTLWIDGTRHVSLEVDGLADSQWKNVTVOAGETYSLVHVGCDLIDSFALDE 177

Db 119 SGLRDEIRYHVIHNGKPRTEAL--PYRMADGQWHRKVALSVASHELLHVDNCRIVERIDP 177

QY 178 PFYELQAKSRMYVAGSARESHFRGLLQNVHLVPNSVEDILSKKCGOQGAENAI 237

Db 178 P--DTNLPPGPNLWLGQRNKHGLFKGIQDKIIF-----MPNGYITQCPNL 223

QY 238 SENTETLRGLPHVTTEYVGPSSERRPEVCERSCELGNNVQELSGHLV-----VNQP 290

Db 224 NHTCPT-----CSDFLSLVQIMDLQELAKWTAKLNYA 257

QY 291 SENLKRYSDNQFLWELIGGPPKTRNKSACWQDREFAENETWV-VDSCTTCTCKKFKTI 349
Db 258 ETRLSQLEN-----CHCEKTCQVSGLLRQDQSWVDGDCRNCCTCKSGAVE 303
QY 350 CHQITCPTATCAS---PSFVEGECC---PSCLSH---VDGE-----381
Db 304 CRRMSCPPLNCSPLSPVHIAGQCKVCRPKCIYGGKVLABGQRILTKSCRECRGGVLVK 363
QY 382 -----EGWSPWAE-----W-TQCSVTGSG-T 401
Db 364 ITEMCPPLNCEKDHILPENQCCRCVRGHNFCAEGPGKCGENSECKNWNNTKATCECKSGYI 423
QY 402 QQRGRS-----CDVTSNTCLGPSIQTRACSLSKDT---RIRODGGWHSWPSWSSCS 450
Db 424 SVQDSAYCEDIDCEAAKMHYCHANTVCVNLPLGLYRCDVCPGYIRVDD-----FSC 475
QY 451 V--TCGVGNTRIRKLCNSPVPMGGKNCCKSGRETAKCQAGAPCIDGRWSWPSWPSACTV 508
Db 476 EHDECGSGQ-----HNCDENAICTNTVQGHSC-----502
QY 509 TCAGGIRERTVCNS---PEPOYGGKACVGDVQERQMCNKRSCP-----VDGCL 554
Db 503 TCRPGYVNGTICRAFCEEGCRYGG-TCVAP-----NKCVCPSGFTGSHCEKIDIBCS 554
QY 555 SN--PCPPGAQCSFPDGSWSCGCPVGFGLNGTH-----CEDLDECALVPDICTFSTSK 606
Db 555 EGIIECHNHSRCVNLPL--GWYHCECRSGFDDGTYSLSGESCIDIDECALRTHTCWNDS- 611
QY 607 VPRCVNTQPFHCLPCP--PRYRGNQVGVGLE-----AAKTEK-QVCEPENCKKOKTH 657
Db 612 --ACINLAGGFDCIL-CPSPGSCSGDCPHEGGLKHNGQVWTLKEDRCVSVC---CKDGKI 664
QY 658 NCHKHAEICVLGHFSDPMYKCECOTGYAGDLGICGSDLDGWNFLNLVLCATNATYHCIK 717
Db 665 FCRTA-----CDCNPSA-DLFCCPE-----CDTRVTSQCLD 696
QY 718 DNCPLHPLNSGOEDFDKDGIGDADDDDDNDGVTDKONCQLLENPR-QADYDKDEVGDRC 776
Db 697 QNGHKLYRSG-----DNWTHSCQCCRLGEVD---CWPLTCPLNLSCEYTAILEGCC 746
QY 777 DNCPIVHNPAQIDTNNEGDACSVDIDGDDVFNERNCPYVNTDQDRTDGDGV-----831
Db 747 PRC-----VSDPLAD---NITYDIRKTC-----LDSYGVSRLSG 778
QY 832 -----GDHCNCPVHNPDQTDVNDLVGDCDNE---DIDDDGHQNN 872
Db 779 SVWTMAGSPCTTC-----KCKNGRVCCSVDFECLQNN 810

RESULT 12

US-09-273-565-34

; Sequence 34, Application US/09273565A

; Patent No. 6166190

; GENERAL INFORMATION:

; APPLICANT: FUJIWARA, TSUTOMU

; APPLICANT: WATANABE, TAKESHI

; APPLICANT: HORIE, MASATO

; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN

; FILE REFERENCE: Q-53599

; CURRENT APPLICATION NUMBER: US/09/273,565A

; CURRENT FILING DATE: 1999-03-22

; EARLIER APPLICATION NUMBER: 09/055,699

; EARLIER FILING DATE: 1998-04-07

; EARLIER APPLICATION NUMBER: 08/820,170

; EARLIER FILING DATE: 1997-03-19

; EARLIER APPLICATION NUMBER: JP 63410/1996

; EARLIER FILING DATE: 1996-03-19

; EARLIER APPLICATION NUMBER: JP 69163/1997

; EARLIER FILING DATE: 1997-03-05

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 34

```

; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-273-565-34

```

| Query Match | 5.9% | Score 388.5 | DB 4 | Length 810 | |
|-----------------------|--------------|--|----------------|------------|---------|
| Best Local Similarity | 20.6% | Pred. No. 8.3e-22 | | | |
| Matches 208 | Conservative | 106 | Mismatches 347 | Indels 349 | Gaps 50 |
| QY | 8 | LALWVW-----PSTQAGH-QDKORTFDLFSISNIRKKTIGAKQFRGPDGPVAPVRFVRFD 61 | | | |
| DB | 5 | LILVWFVCVCTATVVGFGMDPLOMDIVITELDIVNTLTGLVAQVSGMHNASKALFQDIE 64 | | | |
| QY | 62 | -----YIPPVNADLSKTKIMROKEGFLLTAQLKQDKSGRTLLALEGPGLSQRQFEIVS 117 | | | |
| DB | 65 | REIHAAPHVS-----EKLIQLFONKSEFTILATVQOKPSTGCVLSIR--ELEHSYFELES 118 | | | |
| QY | 118 | NGPADYLDLTWYDGTBRHVYSLEVDGLADSQKNWTVQVAGETYSLHWGCDLIDSFALDE 177 | | | |
| DB | 119 | SGURDEIRYHIHNGKRPTEAL-PYRMADQGMHVALVSASHULLHVDCHRIYERVIDP 177 | | | |
| QY | 178 | PFYEHLOAEKSRMYVAKVSGARESHFRGLLQNVHLVFNSEVEDILSKKCGOQOGAEINAI 237 | | | |
| DB | 178 | P--DTNLPPOINLWLGORNKHGLFKGIIDGKIIF-----MPNGYITQCPNL 223 | | | |
| QY | 238 | SENTETLRLGPHVTVTVGSSERRPVCERSCEELGNMVQELSLHLV-----VNQP 290 | | | |
| DB | 224 | NHTCPT-----CSDFLSLVQGINDLQELLAKMTAKLNYA 257 | | | |
| QY | 291 | SENLRKVSNDQFLWELIGGPKTRNMSACWQDGRFAENETW-VDSCTTCTCKKFTI 349 | | | |
| DB | 258 | ETRLSQLEN-----CHCEKTCQVSGLLYRDQSDWYDGHCRNCTCKSGAVE 303 | | | |
| QY | 350 | CHOITCPTAPCAS---PSFYVEGCC---PSLHS---VDGE-----381 | | | |
| DB | 304 | CRMSCPLNCPSDLPVHIAGOCKVCPRKCIYGGKVLAEGORILTKSCRECRGGVLVK 363 | | | |
| QY | 382 | -----EGWSPWAE-----W-TQCSVTGSG-T 401 | | | |
| DB | 364 | ITEMCPPLNCSEKDHILPENQCCRVCRGHNFACBPKGENSECKNNWTKATCECKSGYI 423 | | | |
| QY | 402 | QQRGRS-----CDVTSNLTCLGPTQTRACLSKCDT---RLRQDGGWHSWSPWSSCS 450 | | | |
| DB | 424 | SVQGDSCYCEDIDECRAKMHYCHANTVCVNLPLGYRCDVPGYIRVDD-----PST 475 | | | |
| QY | 451 | V--TCGVGNTRIRLNCNSPVPMGGKKNCKSGRRTKACQACPCPIDGRWSPWSPWACTV 508 | | | |
| DB | 476 | EHDECSGQ-----HNCDENAICTNTVQGHSC-----502 | | | |
| QY | 509 | TCAGGIRERTVNCNS---PEPOYGGKACVDGVQERQMCNKRSCTP-----VDGCL 554 | | | |
| DB | 503 | TKPGYVGNGTICRAFCEBGRYGG-TCVAP-----NKCVCPSGPTGSHCEKXIDECSS 554 | | | |
| QY | 555 | SN--PCFPFGACSFDPGWSGSCSPGVFLGNTH-----CEDLDECALVPDICFSTSK 606 | | | |
| DB | 555 | EGIIETCHNHSRCVNLN-PGWYHCECRSGFDDGTYSLSGESCIDIDECALRTHTCWNDS- 611 | | | |
| QY | 607 | VPRCVNTQPGFHLPCP--PRYRGNQPVGVLE-----AAKTEK-QVCEPENPCKDKTH 657 | | | |
| DB | 612 | --ACINLAGGFDCL-CPSPGSCGDCPHEGGLKHNGQVWTLKEDRCSVCS---CKDGKI 664 | | | |
| QY | 658 | NCHKHAEICVLGHFSDPMYKCEQTVAGGLGIGEDSDLDGWNLVLCATNATYHCIK 717 | | | |
| DB | 665 | FCRRTA-----CDQNPSA-DLFCCPE-----CDTRVTSQCLD 696 | | | |
| QY | 718 | DNCPHLPNSGQEDFDKDIGDADDDDDNDGVTDKDNQCOLLFNPR-QADYDKDEVGDR 776 | | | |
| DB | 697 | QNGHKLYRS-----DNWTHSCQOCCLEGEVD---CWPLTCPLNSCEYTAILEGCC 746 | | | |
| QY | 777 | DNCPYVHNPAQITDNNNGEGACSVDIGDDVFNERNDCNCPYVNTDORDTDGQV-----831 | | | |
| DB | 747 | PRC-----VSDPCLAD---NITYDIRTC-----LDSYGSRLSG 778 | | | |

```

Qy  332  -----GDHCDCNCPYVNPQTOVDNDLVGDQCDNNE-----DIDDDGHQNN 872
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  779  SVWTMAGSPCTTC-----KCKNGRVCVCSDFECLQNN 810

RESULT 13
; Sequence 34, Application US/09565538
; Patent No. 6333404
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBILQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: Q-53599
; CURRENT APPLICATION NUMBER: US/09/565,538
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/273,565
; PRIOR FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-565-538-34

```

| | | |
|----|---|--|
| | Query Match | 5.9%; Score 388.5; DB 4; Length 810; |
| | Best Local Similarity | 20.6%; Pred. No. 8.3e-22; |
| | Matches | Conservative 106; Mismatches 347; Indels 349; Gaps 50; |
| Qy | 8 LALVWV-----PSQAQH-QDKDFTDLFSISNNRKTICAKQPRGDPGPVAPRVRFED | 61 : : : : : : |
| Dd | 5 LILVWFVCVTARTVWGMDPDQLMDIVTELDLVNTLGLVAQVSGMHNASKAFPLFODIE | 64 : : : : : : |
| Qy | 62 ---YIPVNADLLSKTKMKROKGEFFLAQLKDQCKSRGTLLALREGLSORQRFEIVS | 117 : : : : : : : : : : : : : : : : : |
| Dd | 65 REIIAAAPHVS----EKLIQIOLFONKSEFIILTATVOOKPSTSGVILSIR--ELEHSYFLES | 118 : : : : : : : : : : : : : : : : : |
| Qy | 118 NGPADTLDLTWIDGTRHVVSLDEDVLGADSQWNVTVOAGETYSLVHGCDLIDSFALDE | 177 : : : : : : : : : : : : : : : : : |
| Dd | 119 SGLRDEIRYHYIHNGKPRTREAL-PYRMADQWHKVVALSVASHLHLLHVDCNRIVERIDP | 177 : : : : : : : : : : : : : : : : : |
| Qy | 178 PFYEHLQAEKSMYAVKGSARESHFRLGLQNVLHVPNSVEDILSKGCQOGGAELNAI | 237 : : : : : : : : : : : : : : : : : : : : : |
| Dd | 178 P--DTNLPPGTLNWLGNORKHGLFKGIIODGKIIF-----MPNGYITOCPNL | 223 : : : : : : : : : : : : : : : : : : : : : |
| Qy | 238 SENTETLRGLPHVITYEYVGSSRRPEVCERSCEELNMVQELSGLHVL-----VNQP | 290 : : : : : : : : : : : : : : : : : : : : : |
| Dd | 224 NHTCPT-----CSDFLSLVQOIMDLQELLAKMTAKLNYYA | 257 : : : : : : : : : : : : : : : : : : : : : |
| Qy | 291 SENLKRYSDNQFLWELIGPPKPTRNMSACWQDGRFAENETWV-VDSCTTCCTCKFKFTI | 349 : : : : : : : : : : : : : : : : : |
| Dd | 258 ETRLSOLE-----CHCEKTCQVSGLLTYRDQDSWVDGDHCRNCTCKSGAVE | 303 : : : : : : : : : : : : : : : : : |
| Qy | 350 CHQITCPPATCAS---PSFVEGEC-----PSCLHUS-----VDGE----- | 381 : : : : : : : : : : : : : : : : : : : : : |
| Dd | 304 CRRMSCPPLNCSPDSPLVHIATGCCVKCRPKCIYGGVKGVALAEGORILTKSCRECRGGVLYK | 363 : : : : : : : : : : : : : : : : : |
| Qy | 382 -----PGWSFWAE-----W-TCCSYTCGSG-T | 401 : : : : : : : : : : : : : : : : : |
| Dd | 364 ITEMCPPLNSEKHILPENQCRCVRGHNFCAEPKCGENSECKNNWTATCEKSGYI | 423 : : : : : : : : : : : : : : : : : |

```
QY 402 QORGRS-----CDVTSNTCLGSIOTRACSLSKDT---RIRODGGHSHWSPWSSCS 450
Db 424 SVQGSAYCEDIDCAKMHVCHANTVCVNLPLGLYRCDVPGYIRVDD-----FSCT 475
QY 451 V--TCGVGNITRILCHNSPVQMGKCKGSGRETRKACQCAPCPIDGRWSPWSPWACTV 508
Db 476 EHDECGSQ-----HNCDENAICTNTVQGHSC-----502
QY 509 TCAGGIRTRVCHNS---PEPOYGGKACVGDVQERQMCNKRSCP-----VDGCL 554
Db 503 TKCPGYVNGNTICRAFCEGCRYGG--TCVAP-----NKCVCPSGFTGSHCEKIDEC 554
QY 555 SN--PCFPAGACSSFPDGSWCGSCPVGFLNGTH-----CEDLDECALVPDLCFSTSK 606
Db 555 EGIIECHNSHCYNLP--GWYHCECRSGFHDGYSLSGSCDIDICALRTHTCWDS- 611
QY 607 VPRCVNTQPGFHLPLCP--PRYRGNQPVGVGLE-----AAKTEK-QVCPEPNCKDKTH 657
Db 612 --ACINLAGGFDCL-CFSGPSCGDCPHEGGLKHNGQVWTLKEDRCVCS----CKDGI 664
QY 658 NCHKHAECIYLGHSDPMYCECQTGAGDGLIGEDSDLDGWNLVNLCATNATYCIK 717
Db 665 FCRRTA-----CDQNPASA-DLFCCPE-----CDTRVTSQCLD 696
QY 718 DNCPLHNSGOEDFDKDGIGDADCDDDDDNDGVTDEKDNCLLFNPR-QADYDKDEVDGRC 776
Db 697 QNGHLYRSG-----DNWTHSCQOCRLGEVD----CHPLCPNLSCEYTAILEGCC 746
QY 777 DNPYVHNPAQIDTNNNEGSDVSDIGDDVFNERNRNCPPYVNTDQDRTDGGV----- 831
Db 747 PRC-----VSDPCLAD--NITYDIRKTC-----LDSYGVSRSLG 778
QY 832 -----GDHCDNCLVHNPDQTDVDNDLVGDQCDNNE---DIDDDGHQNN 872
Db 779 SVWTMAGSPCTC-----CKNKRVCVSDPECLQNN 810

RESULT 14
US-08-820-170A-37
; Sequence 37, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELETYPE: 6491103
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-820-170A-37
```

```
Query Match 5.5%; Score 361.5; DB 2; Length 816;
Best Local Similarity 20.1%; Pred. No. 1.1e-19;
Matches 204; Conservative 126; Mismatches 354; Indels 333; Gaps 49;

QY 2 VRLVLVLLALWVWSTQAGHQDKDTTDFLFSISINRRTIGAKQRPDPGPVAYRFRVFD 61
Db 19 VW-----GLGVDFSLQ-----IDVLELGESTGTGVRQVPLHNGTKAFLFQDTP 64
QY 62 YIPVNNADLSKITKIMROKEGFLTAQLKODGKSRGRTLLALEGPGLSQROFELVSNP 121
Db 65 RSIAKATAAEQFQKLRNKHETILVTLKQTHLNSVILSIH--HLDHYLELESSGHR 122
QY 122 DTLDTLWIDGTRHVSLEVDGLADSQMKNVTVOVAGETYSLHVGCDLIDSFALDEPFE 181
Db 123 NEVRLHYRSGSRHPHTEVPFVILADDKWHKLSAISASHLILHIDCNKI-----YE 173
QY 182 HLQAEKS-----RMVYAKGSARESHFRGLLQNVHLVFNSEVEDILSKGCGQCGAEI 234
Db 174 RVVEKPSDPLGTTFWLGQRNNAHGYFKGIMQDVQLLV-----MPQGFIAQC 221
QY 235 NAISENTETLRLGPHVTTEYVGPSSERRRPEVCERSCELGNNVQBELSGLHVLVNPSEN 294
Db 222 PDNRTCTP-----CNDFHGLVQKIMELQDILAKTSAKL 255
QY 295 -----KRVSNQNFQELWELIGPPKTRNMSACWQGRFFAENETWVVDSCITCTCKKFTIC 350
Db 256 SRAEQRMRLDQCYCE-----RTCTMKGTTRYEFESW-IDGCKNCTCLNGTIOC 303
QY 351 HOITCPATC---ASPSFVEGECPCSLHSDGEGHSPAENWTOCSVTCSGSGTOQRGRS 407
Db 304 ETLICPNPDCPLKSAIAYVDKCC-----KEKSIK---QFGRT 340
QY 408 -CDVTSNTCLGPSIOTRACSLSKDTRIRQDGGHSHWSPWSSCSVTGCVGNITRILCNS 466
Db 341 YFEGERTNYVSSS---GVCVLYECKDQ-----TMKLV-- 369
QY 467 PVQMGKNCCKSGRETKACQAGAPCIDGRWSPWSPWSPWACTVTCAGGIRTRVCNPEP 526
Db 370 -----ESSGCPALDC-----PESHOITLHSHCC----- 392
QY 527 QYGGKACVGDVQBRQMCNKRSCPVDGCLSNPCFPAGACSSFPDGSWCGSCPVGFLG--- 583
Db 393 ---KVCCKG---YDFCSE-----HNCMENSICRNLDRA-VC-SCRDGFRALRE 433
QY 584 NGTHCEDLDECALVPDLCFSTSKVPRCVNTQPGFHLCPCPPRYRGNQPVGVGLEAAKTEK 643
Db 434 DNAYCEDIDCAEGRHYCRENT---MCVNTPGSFMCI-CKTGY-----IRIDD 477
QY 644 QVCEPENPKDKTHNCHKHAECLYL--GHFSDPMYKCECOTGVAGDGLIGEDSDLDGWP 701
Db 478 YSCTEHDECITNQHNCDENALCFNTVGGH-----NCVCRPGYTGNGTTC-KAFCKGCR 530
QY 702 NLMLVCAATNATYHCIKDNCPLHNSGOEDFDKDGIGDADCDDDDD--NDGVT--DEKDN- 756
Db 531 NNGGACIAANV---CA---CP-----QGTGTPSCETDIDECSDGFVQCDSRANCI 573
QY 757 -----QLLFNPROADYDKDEVDGRDNCPCPVHNPADITDNNNEG----- 796
Db 574 NLPQWYHCECRDGYVHDNGMFSPS-----GESCE-----DIDECGTGRHS 613
QY 797 --DACSVDDIDGDDVFNERNRNCPPYVNTDQDRTDGDGVGDH-----CDNC 838
Db 614 ANDTICFNLDDG--YDCR--CPHGKNC-TGDCIHDGKVKHNGQIWLLENDRCSVCSCNG 668
QY 839 PLVHNPDQTDVNDLVG-----DQCD-----NNEDIDDDHQNNDNCPPYISANQADHRRD 891
Db 669 FVCMRRMVCDCENPTVDLFCCPECDPRLSSOCLHQNGETLYNSGDTWVQNCQCC---RCL 725
```


QY 892 QGDA-----C-DPDDNDGVPDRDRNCRVFNPDQEDLDGDRGDIKDDFDNDNI 941
Db 726 QGEVDCWPLPCPDVECFEFSILPENECCPRCVTDPCQADTI---RNDITKTCLDEMNV 779

RESULT 15

US-09-055-699-37
Sequence 37, Application US/09055699
Patent No. 6005088

GENERAL INFORMATION:

APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESS: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-055-699-37

Query Match 5.5%; Score 361.5; DB 3; Length 816;
Best Local Similarity 20.1%; Pred. No. 1.1e-19; Mismatches 354; Indels 333; Gaps 49;
Matches 204; Conservative 126

QY 2 VNRLLVLLWVWPSTQAGHQDQDTTDFLFSININKRTIGAKQFRGPDGVPAYRFRVD 61
Db 19 VW-----GLGVDFSLQ-----IDVLTELELGESTTGVQVPCVCLHNGTKAFLFQDTP 64
QY 62 YIPVNAVDDLKTKTKMROKEGFFLTAQKQDGKSGRTLLALEGPGLSORQFEIVSNGPA 121
Db 65 RSIKASTATAEQPFQKLRNKHFEITLVTLKQTHLNSGVLSIH--HLDHRYLELESSGHR 122
QY 122 DTLDLYTWIDGTRHVVSLVDVGLDSQWKNVTVOVAGETYSLVHVGCDLIDSFALDEPFYE 181
Db 123 NEVRLYRSGSHRPHTEVPYILADKWHKLSAISASHILHIDCNKI-----YE 173
QY 182 HLOAKES-----RMVAKGSARESHFRLLQNVHLVFNSEVEDILSKKQCOQOGAEI 234
Db 174 RVVEKPTDPLGTTFEVLGORNNAHGVFKGIMQDVQLV-----MPQGFIAQC 221
QY 235 NAISENFETLRGLPHVTTEYVGPSSRRPVCERSCVEELGNMVQELSGHLVLPQSEN 294
Db 222 PDLNRTCT-----CNDFHGLVQKIMELQDILAKTSAKL 255
QY 295 -----KRVSDNDNQFLWELIGPPKPNKNSACWDQGRFFAENETWVDSCTTCTCKKFKTIC 350

Db 256 SRAEQMNRDLQCYCB-----RCTMKGTYYREFESW-IDGCKNCTCLNGTIQC 303
QY 351 HQITCPPATC---ASPSFVEGECPCSLHSVDGEGSPMAEWTCQSVTCGSGTQGRGRS 407
Db 304 ETLICPNPDCPLKSALAYVDGKC-----KECKSIC---QFGRT 340
QY 408 -CDVTSNTCLGPSIQTRACSLSKCDTRIQRDGGGSHWSPSSCVTCGVGNITRILCNS 466
Db 341 YFEGERNVYSSS---GVCVLYECKDQ-----TMKLV--- 369
QY 467 PVPQMGKCKGSGRETAKACQCAPCPIDGRWSPWSPWACTVTCAGGIRERTRVCNSPEP 526
Db 370 -----ESSGCPALDC-----PESHQITLSHSCC----- 392
QY 527 QYGGKACVGDVQERQCNKRSCPVDGCLNPFPGACQSFDPGWSWSSCSCPVGFLG--- 583
Db 393 ---KVCKG---YDFCSE-----HNCMENSICRNLNDRA-VC-SCRDGFRALRE 433
QY 584 NGTHCEDLDECALVPDICEFSTSKVPRCVNTQGFHCLPCPPRYRGVQVGVGLEAAKTEK 643
Db 434 DNAYCEDIDECAGRHRYCRENT---NCVATPGSFMCI-CKTGY-----IRIDD 477
QY 644 QVCEPENCKDKTHNCHKHAECIYL--GHFSDPMYKCEQOTGYAGDGLICGEDSDLDGWP 701
Db 478 YSCTEHEDEITNOHNCDENALCENTVGGH-----NCVCKPGYTGNGTTC-KAFCKDCCR 530
QY 702 NLNLVATNATYHCKNDNCPHLPNSQEDFDKDGIGDADDDDD---NDGVT---DEKDN- 756
Db 531 NGGACIAANV---CA---CP-----QGFTGPSCETDIDECSDGFVQCDSRANCI 573
QY 757 -----QLLFNPRQADYDKDEVDRCDNCPYVHNPAQIDTNNNEG--- 796
Db 574 NLPGWYHCECRDGYHDNGMFSPS-----GESCE-----DIDEGTGRHSC 613
QY 797 --DAGSVDDIDGDDVFNERNDCPVYVNTDQDRTDGDGVGDH-----CDNC 838
Db 614 ANDTICFNLDDG--YDCR--CPHGKNC-TGDCIHDKVKHNGIWIWLENDRCSCVSCQNG 668
QY 839 PLVHNPDPQTDVNDLVG-----DQCD---NNEDIDDDGHQNNQDNCPYISNANQADHRDG 891
Db 669 FVMCRRMVDCENPTVDLFCCECDPRLSSQCLHQNGETLYNSGDTWVQNCQOC---RCL 725
QY 892 QGDA-----C-DPDDNDGVPDRDRNCRVFNPDQEDLDGDRGDIKDDFDNDNI 941
Db 726 QGEVDCWPLPCPDVECFEFSILPENECCPRCVTDPCQADTI---RNDITKTCLDEMNV 779

Search completed: August 9, 2002, 10:01:36
Job time: 162 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 09:59:24 ; Search time 29.54 Seconds
(without alignments)
3812.346 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWVPSSTQAGH.....FVFSQEMVYFSLKAYECRDI 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 6570 | 99.5 | 1172 | 1 TSHUP2 | thrombospondin 2 p |
| 2 | 5920 | 89.6 | 1172 | 2 A42587 | thrombospondin 2 p |
| 3 | 5481 | 83.0 | 1178 | 1 A39804 | thrombospondin pre |
| 4 | 4228 | 64.0 | 1170 | 1 TSHUP1 | thrombospondin 1 p |
| 5 | 4189.5 | 63.4 | 1170 | 2 A40558 | thrombospondin 1 p |
| 6 | 1968 | 29.8 | 755 | 2 A44315 | cartilage oligomer |
| 7 | 1946.5 | 29.5 | 956 | 2 A57121 | thrombospondin 3 p |
| 8 | 1936.5 | 29.3 | 956 | 1 A46016 | thrombospondin 3 - |
| 9 | 1899 | 28.8 | 955 | 2 A45441 | thrombospondin 4 - |
| 10 | 1890 | 28.6 | 961 | 1 TSHUP4 | thrombospondin 4 p |
| 11 | 1480 | 22.4 | 439 | 2 B44315 | cartilage oligomer |
| 12 | 580 | 8.8 | 128 | 2 S57955 | thrombospondin 2 - |
| 13 | 439.5 | 6.7 | 437 | 2 S05478 | properdin - mouse |
| 14 | 429 | 6.5 | 1444 | 2 T18856 | angiogenesis inhib |
| 15 | 424 | 6.4 | 1584 | 2 T00026 | brain-specific ang |
| 16 | 418.5 | 6.3 | 810 | 2 T10756 | Nel-homolog protei |
| 17 | 418 | 6.3 | 984 | 2 T00326 | hypothetical prote |
| 18 | 418 | 6.3 | 1522 | 2 T00028 | brain-specific ang |
| 19 | 416 | 6.3 | 469 | 1 S29126 | properdin precurs |
| 20 | 415 | 6.3 | 1074 | 2 JCS928 | semaphorin F precu |
| 21 | 405 | 6.1 | 957 | 2 T15976 | hypothetical prote |
| 22 | 399 | 6.0 | 788 | 2 T25061 | hypothetical prote |
| 23 | 384.5 | 5.8 | 1572 | 2 T00027 | brain-specific ang |
| 24 | 381 | 5.8 | 229 | 2 S57957 | thrombospondin 1 - |
| 25 | 343.5 | 5.2 | 2531 | 2 A46019 | Notch-1 protein - |
| 26 | 340.5 | 5.2 | 2531 | 2 S18188 | notch protein homo |
| 27 | 339.5 | 5.1 | 835 | 2 JP0076 | nel protein - chlc |
| 28 | 329.5 | 5.0 | 807 | 2 A38152 | F-spondin - rat |
| 29 | 329.5 | 5.0 | 2524 | 2 A35844 | Xotch protein - Af |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 322 | 4.9 | 2471 | 2 | A49128 | cell-fate determin |
| 31 | 318.5 | 4.8 | 2531 | 2 | T31070 | notch homolog - se |
| 32 | 317 | 4.8 | 1360 | 2 | T33922 | hypothetical prote |
| 33 | 312.5 | 4.7 | 861 | 2 | A48825 | Notch homolog Motc |
| 34 | 311.5 | 4.7 | 2352 | 2 | T30201 | Notch homolog prot |
| 35 | 308 | 4.7 | 803 | 2 | A47723 | F-spondin precurs |
| 36 | 307 | 4.6 | 2871 | 2 | A55624 | fibrillin-1 precu |
| 37 | 305.5 | 4.6 | 2703 | 1 | A24420 | notch protein - fr |
| 38 | 302 | 4.6 | 2555 | 2 | A40043 | notch protein homo |
| 39 | 300.5 | 4.5 | 2437 | 2 | S42612 | transmembrane prot |
| 40 | 300 | 4.5 | 3002 | 2 | A47221 | fibrillin 1 precu |
| 41 | 297 | 4.5 | 2871 | 2 | A55567 | fibrillin 1 - bovi |
| 42 | 296 | 4.5 | 2139 | 2 | A35672 | crumbs protein - f |
| 43 | 295 | 4.5 | 712 | 2 | A45638 | immunodominant mic |
| 44 | 292 | 4.4 | 1558 | 2 | C89114 | protein C37C3.6a l |
| 45 | 292 | 4.4 | 2167 | 2 | T34395 | hypothetical prote |

ALIGNMENTS

RESULT 1

TSHUP2
thrombospondin 2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: A47379; A42173
R:LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993
A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA:
A:Reference number: A47379; MUID:94010892
A:Accession: A47379
A:Molecule type: mRNA
A:Residues: 1-1172 <LAB>
A:Cross-references: GB:L12350; NID:9307505; PIDN:AAA03703.1; PID:9307506
R:LaBell, T.L.; Milewicz, D.J.; Distèche, C.M.; Byers, P.H.
Genomics 12, 421-429, 1992
A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expression
A:Reference number: A42173; MUID:92217961
A:Accession: A42173
A:Molecule type: mRNA
A:Residues: 560-1172 <LA2>
A:Cross-references: GB:M81339
A:Experimental source: fibroblast
A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)
C:Genetics:
A:Gene: GDB:THBS2; TSP2
A:Cross-references: GDB:128789; OMIM:188061
A:Map position: 6q27-6q27
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregat
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; ti
F:18/Domain: signal sequence #status predicted <SIG>
F:19-1172/Product: thrombospondin 2 #status predicted <MAT>
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>
F:928-930/Region: cell attachment (R-G-D) motif
F:151-316-330-457-584-710-1069/Binding site: carbohydrate (Asn) (covalent) #status F
F:167-226/Disulfide bonds: #status predicted
F:266,270/Disulfide bonds: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 99.5%; Score 6570; DB 1; Length 1172;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 MWRLVLLALWVWPSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGVPAYRFRVF 60
Db 1 MWRLVLLALWVWPSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGVPAYRFRVF 60
QY 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
Db 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
QY 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVGCDLIDSFALDEFFY 180
Db 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVGCDLIDSFALDEFFY 180
QY 181 EHLQAEKSMYVAKGSAHRESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
Db 181 EHLQAEKSMYVAKGSAHRESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
QY 241 TETLRLGHVHTTEYVGPSSERRPEVCERSCBELGNMVOELSLHLVNVQPSLENKRVNSD 300
Db 241 TETLRLGHVHTTEYVGPSSERRPEVCERSCBELGNMVOELSLHLVNVQPSLENKRVNSD 300
QY 301 NOFLWELIGGPPKTRNMSACWDGFRFAENETWVDSCTTCTCKKFKTICHOITCPPATC 360
Db 301 NOFLWELIGGPPKTRNMSACWDGFRFAENETWVDSCTTCTCKKFKTICHOITCPPATC 360
QY 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
Db 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
QY 421 QTRACSLSKCDTRIQDGGSHWSPSSCVTCGVNTRIRLNCSPVPQMGKNCKSG 480
Db 421 QTRACSLSKCDTRIQDGGSHWSPSSCVTCGVNTRIRLNCSPVPQMGKNCKSG 480
QY 481 RETKACQAPCIDGRWSPWSPWACTVTCAGGIRERTVNCNPPPOYGGKACVGDQVER 540
Db 481 RETKACQAPCIDGRWSPWSPWACTVTCAGGIRERTVNCNPPPOYGGKACVGDQVER 540
QY 541 QMCNKRSPVGCGLSNPCFPAQCSFPGDWSGSCSPVGFIGNTHCEDLDECALVPDI 600
Db 541 QMCNKRSPVGCGLSNPCFPAQCSFPGDWSGSCSPVGFIGNTHCEDLDECALVPDI 600
QY 601 CFSTSKVPCRCVNTQPGFHCPCPPRYRGNQPVGVGLEAKTEKQVCEPENCKDKTHCH 660
Db 601 CFSTSKVPCRCVNTQPGFHCPCPPRYRGNQPVGVGLEAKTEKQVCEPENCKDKTHCH 660
QY 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCATNATYHCIDNC 720
Db 661 KHAECIYLGHSFSDPMYKCECOTGYAGDGLICGEDSLDGLWPNLNLVCATNATYHCIDNC 720
QY 721 PHLPNSGQEDFKDGIIGDADDDNDGVTDEKDNCLLFPNRQADYDKDEVDGRDCNCP 780
Db 721 PHLPNSGQEDFKDGIIGDADDDNDGVTDEKDNCLLFPNRQADYDKDEVDGRDCNCP 780
QY 781 VYHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRTDGDGVDHDCNCP 840
Db 781 VYHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRTDGDGVDHDCNCP 840
QY 841 VHNPDQTDVNDLVGDQDNNEDIDDDGHQNNODNCPYISNANQADHSDGOGDADCPDD 900
Db 841 VHNPDQTDVNDLVGDQDNNEDIDDDGHQNNODNCPYISNANQADHSDGOGDADCPDD 900
QY 901 DNDGVPDRDNCRLVFNPDQEDLDGGRGDIKDDFDNDNIPDIDDDVCPENNAISETDFR 960
Db 901 DNDGVPDRDNCRLVFNPDQEDLDGGRGDIKDDFDNDNIPDIDDDVCPENNAISETDFR 960
QY 961 NFQMPVLPDPKGTQIDPNWVIRHQKELVQVANSDPGLAVGDFSGVDFSGTFVYNTDR 1020
Db 961 NFQMPVLPDPKGTQIDPNWVIRHQKELVQVANSDPGLAVGDFSGVDFSGTFVYNTDR 1020
QY 1021 DDDYAGVFVGQSSSRFYVMKQVQTYWEDQPTFRAYGSGVSLKVVNSTGTGEHLRN 1080
Db 1021 DDDYAGVFVGQSSSRFYVMKQVQTYWEDQPTFRAYGSGVSLKVVNSTGTGEHLRN 1080
QY 1081 ALWHTGNTPGQVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
```

```
Db 1081 ALWHTGNTPGQVRTLWHDPRNIGWKDYTAIRWHLTHRPKTYGIRVLVHEGKQVMAADSGPI 1140
QY 1141 YDOTYAGRLGLFVFSQEMVYFSDLYKECRDI 1172
Db 1141 YDOTYAGRLGLFVFSQEMVYFSDLYKECRDI 1172

RESULT 2
thrombospondin 2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C:Accession: A42587; A39851
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ct
A:Reference number: A42587; MUID:92147683
A:Accession: A42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1172 <LAH>
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.
J. Biol. Chem. 266, 12821-12824, 1991
A:Title: A second, expressed thrombospondin gene (thbs2) exists in the mouse genome.
A:Reference number: A39851; MUID:91302287
A:Accession: A39851
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-873 <BOR>
A:Cross-references: GB:M64866; NID:g201994; PIDN:AAA0432.1; PID:g201995
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
E:Keywords: calcium binding; glycoprotein
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F:553-588/Domain: EGF homology <EGF1>
F:652-691/Domain: EGF homology <EGF>

Query Match 89.6%; Score 5920; DB 2; Length 1172;
Best Local Similarity 88.5%; Pred. No. 0;
Matches 1036; Conservative 57; Mismatches 78; Indels 0; Gaps 0;
QY 1 MWRLVLLALWVWPSTQAGHODKDTTDLFSISINIRKTIAGAKQRPDPGVPAYRFRVF 60
Db 1 MLWALLALGIGPRASAGDHVKDTSFDLFSISINIRKTIAGAKQRPDPGVPATFRFRVF 60
QY 61 DYIPVNAADDLSKITKIMROKEGFFLTALQKODGKSRGTLALLEGPGISQRFVSNRP 120
Db 61 DYIPVNTDDLNRIVKLARRKEGFFLTALQKDRKSRTLLVLEGPGTSQRFVSNRP 120
QY 121 ADTLDLTWIDGTRHVSLEVDGLADSOQKNTVQVAGETYSLVHVGCDLIDSFALDEFFY 180
Db 121 GDTLDLNTWVEGNQHTNFDVGLADSOQKNTVQVADTSYLVYVGCGLIDSVTLEEFY 180
QY 181 EHLQAEKSMYVAKGSAHRESHFRGLLQNVHLVFNSEVEDILSKKCGQCGQGAENAISEN 240
Db 181 EQLVDESRMTYVAKGSAHRESHFRGLLQNVHLVFNSEVEDILSKKCGQSAQEAVENTISEH 240
QY 241 TETLRLGHVHTTEYVGPSSERRPEVCERSCBELGNMVOELSLHLVNVQPSLENKRVNSD 300
Db 241 TETLHLSPHITDLVVOGVEKAQEVCTHSCBELSNMNLSELGLHVMVNLKNLERSVD 300
QY 301 NOFLWELIGGPPKTRNMSACWDGFRFAENETWVDSCTTCTCKKFKTICHOITCPPATC 360
Db 301 NOFLLELIGGLKTRNMSACVQEGRIFAENETWVDSCTTCTCKKFKTVCHOITCSPATC 360
QY 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
Db 361 ASPSFVEGECPCSLHSDVGEKSWPWAETQCSVTCSGSGTQQRGRSCDVTSTNTCLGPSI 420
```

Db 361 ANPSFVEGCCPSCSHSADSDGSPWAEWTECSVTGCGSTQGRSCDVTNTCLGPSI 420
QY 421 QTRACSLSKCDTRI RDGGSWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGSG 480
Db 421 QTRFCSLKGKCDTRI RQNGGWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGSG 480
QY 481 RETKACQAGCPIDGRWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGSG 540
Db 481 RETKPCQORDPCPIDGRWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGSG 540
QY 541 QMCKNSRCPVDGCLSNPCFPAGCSCFPDGSWSGCGSPVGLNGTHCEDLDCAVLPDI 600
Db 541 QMCKNSRCPIDGCLSNPCFPAGCSCFPDGSWSGCGSPVGLNGTHCEDLDCAVLPDI 600
QY 601 CFSTSKVPRCVNTQPGFCHLPCPPRYRGNOVPGVGLAETKQVCEPENCKDKTHNCH 660
Db 601 CFSTNKAPRCVNTNPGFCHLPCPPRYRGNOVPGVGLAETKQVCEPENCKDKTHSCH 660
QY 661 KHAECIYLGHFSDPMYKCECQIAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 720
Db 661 KHAECIYLGHFSDPMYKCECQIAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 720
QY 721 PHLPNSQEDFDRDGIGDADCDNDNDGVTDERKNCQLLFPNPRQADYDKDEVDGDRDNC 780
Db 721 PKLPNSQEDFDRDGIGDADCDNDNDGVTDERKNCQLLFPNPRQADYDKDEVDGDRDNC 780
QY 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYVNTDQDRTDGGVGHDCNCPL 840
Db 781 YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYVNTDQDRTDGGVGHDCNCPL 840
QY 841 VHPNDQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANQADHDDRGQDGDADPD 900
Db 841 MHPNDQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANQADHDDRGQDGDADPD 900
QY 901 DNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 960
Db 901 DNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 960
QY 961 NFQWPLDPKGTQIDPNWVIRHQGKELVQTANSDFGIAVGFDFGSGVDFSGTFYVNTDR 1020
Db 961 NFQWPLDPKGTQIDPNWVIRHQGKELVQTANSDFGIAVGFDFGSGVDFSGTFYVNTDR 1020
QY 1021 DDDYAGVFYQSSRFYVMVQVYQTYWEDQTRAYGSGYSLKVVNSTTGTGHELRN 1080
Db 1021 DDDYAGVFYQSSRFYVMVQVYQTYWEDQTRAYGSGYSLKVVNSTTGTGHELRN 1080
QY 1081 ALMHTGNTPGQVRLMHPDRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVYADSGPI 1140
Db 1081 ALMHTGNTPGQVRLMHPDRNIGWKDYATYRWHLTHRPKTYIRVLVHEGKQVYADSGPI 1140
QY 1141 YDQTYAGRLGLVFSQEMVYFSDLYKECRD 1171
Db 1141 YDQTYAGRLGLVFSQEMVYFSDLYKECRD 1171

RESULT 3

A39804

thrombospondin precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

R:Lawler, J.; Duquette, M.; Ferro, P.

J. Biol. Chem. 266, 8039-8043, 1991

A:Title: Cloning and sequencing of chicken thrombospondin.

A:Reference number: A39804; MUID:91217026

A:Accession: A39804

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1178 <LAW>

A:Cross-references: GB:M60853; NID:g212763; PIDN:AAA51437.1; PID:g212764

C:Species: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vc

F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>

F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>

F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:658-697/Domain: EGF homology <EGF>

Query Match 83.0%; Score 5481; DB 1; Length 1178;

Best Local Similarity 79.7%; Pred. No. 0;

Matches 934; Conservative 110; Mismatches 126; Indels 2; Gaps 2;

QY 1 MWRLVLLALWMPSTQAGHQDKDTTDFLSISINIRKKTIGAKOFKRGDPGVPAYRFRVF 60

Db 7 LLWLAVFITLWSSDAQDDAKEENTFDLLQISINIRKKTIGAKLFRGPDPAIPAYRFRVF 66

QY 61 DYTPPNADSLKTIIMRQKEGFFLTAQLKQDKGRSGTLLEGGLSQRQFEIVSNGP 120

Db 67 DHIPPFKEPKLKVILIRQNEGFILSATLRQDRQSRGTILLEGGLSQRQFEIVSNGR 126

QY 121 ADTLDLTYWIDGRHVRVSLVDGLADSQMKNVTQVAGETYSLVHVGDLDSFALDPFFY 180

Db 127 ANTLDLIYWDGPNVLSLVDGLADSQMKNVTQVAGETYSLVHVGDLDSFALDPFFY 186

QY 181 EHLQAKSRMYVAKGSAHSHFRLGLLQNVHLVENSVEDILSKKGGQGGQGAENAISEN 240

Db 187 EQLKAENSRMYVAKGSIENHFRGLLQNIHLIEDTSIEDVLRKKGQSRQSTEVNTINES 246

QY 241 TETLRGLGPHVTTEYVGPSSERRPERCEELGNMVOELSLGLHVLVNQPSNLRKVSND 300

Db 247 TELHLSHPAVTTEYVGEKTEKKAFCDRSCEELGTFTLTGLRIVVNNADNLQKVSEE 306

QY 301 NQFLWELIGPPKKT-RNMSACWQDGRFFFAENETWYVDSCTCTCKKFKTKCHQITCPPAT 359

Db 307 NQIMWELI-GPNKTLANQSVCMQDGRVADSESWIVDSCTCKTCQDSKIVCHOITCPPVS 365

QY 360 CASPSFVEGECPCSLHSVDGEGSWPBAWTCOSVTCGSGTQGRSCDVTNTCLGPS 419

Db 366 CADPSFIEGECPCSVCSHSDSEEGSWPBDWTKCSVTGSGTQGRSCDVTNTCLGPS 425

QY 420 IQTRACSLSKCDTRI RDGGSWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGS 479

Db 426 IQTRACSLSKCDTRI RDGGSWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGS 485

QY 480 GRETAKCQAGPACPIDGRWSHSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGS 539

Db 486 GRETEKCEKAPCPVNGQWSPWSSCSVTGCGVGNITRILCNLSPVPMQMGKCKGS 545

QY 540 RQMCNRSRCPVDCCLSNPCFPAGCSCFPDGSWSGCGSPVGLNGTHCEDLDCAVLPD 599

Db 546 HDMCNKRDCPIDGCLSNPCFPAGCSCFPDGSWSGCGSPVGLNGTHCEDLDCAVLPD 605

QY 600 ICFSTSKVPRCVNTQPGFCHLPCPPRYRGNOVPGVGLAETKQVCEPENCKDKTHNC 659

Db 606 VCFKVNQVIRCVNTNPGFCHLPCPPRYRGNOVPGVGLAETKQVCEPENCKDKTHNC 665

QY 660 HKHAECIYLGHFSDPMYKCECQIAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 719

Db 666 HKHAECIYLGHFSDPMYKCECQIAGDGLICGEDSDLDGWPNNLVCATNATYHCKDKNC 725

QY 720 CPHLPNSQEDFDRDGIGDADCDNDNDGVTDERKNCQLLFPNPRQADYDKDEVDGDRDNC 779

Db 726 CPHLPNSQEDFDRDGIGDADCDNDNDGVTDERKNCQLLFPNPRQADYDKDEVDGDRDNC 785

QY 780 PYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYVNTDQDRTDGGVGHDCNC 839

Db 786 PYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYVNTDQDRTDGGVGHDCNC 845

QY 840 LVHNPQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANQADHDDRGQDGDADPD 899

Db 846 LVHNPQTDVNDLVGQDCNNEDIDDDGHONNODNCPYISNANQADHDDRGQDGDADPD 905

QY 900 DDNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 959

Db 906 DDNDGVPDRDRNCLVFPNPDQEDLDGGRGDIKDDFDNDNIPDIDVCPENNAISETDF 965

QY 960 RNFQVPLDLPKGTQTIDPNWVIRHQKELVQVANSOPGIAVGDFGSGVDFSGTFVYNTD 1019
 Db 966 RKFQVPLDLPKGTQADPNWVIRHQKELVQVANSOPGIAVGDFGSGVDFSGTFVYNTD 1025
 QY 1020 RDDYAGVFGVQSSRFVVMKQVOTQYWEOPPRAYGSGVSLKVNSTGTGGEHLR 1079
 Db 1026 RDDYAGVFGVQSSRFVVMKQVOTQYWEOPPRAYGSGVSLKVNSTGTGGEHLR 1085
 QY 1080 NALWHTGNTPGQVRLWHDPRNIGMKDYATYRWHLTHRPKTYIRVLVHVGQVMDASGP 1139
 Db 1086 NALWHTGNTPGQVRLWHDPRNIGMKDYATYRWHLTHRPKTYIRVLVHVGQVMDASGP 1145
 QY 1140 IYDQYAGRLGLFVFSQEMVYFSDLKYCRD 1171
 Db 1146 IYDQYAGRLGLFVFSQEMVYFSDLKYCRD 1177

RESULT 4
 TSHUP1
 thrombospondin 1 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
 C:Accession: A26155; A34274; A30140; A25812; A05172; A42927
 R:Lawler, J.; Hynes, R.O.
 J. Cell Biol. 103, 1635-1648, 1986
 A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple C
 A:Reference number: A26155; MUID:87057617
 A:Accession: A26155
 A:Molecule type: mRNA
 A:Residues: 1-1170 <LAW>
 A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 J. Lahterty, C.D.; Gierman, T.M.; Dixit, V.M.
 J. Biol. Chem. 264, 11222-11227, 1989
 A:Title: Characterization of the promoter region of the human thrombospondin gene. DNA s
 A:Reference number: A34274; MUID:89291870
 A:Accession: A34274
 A:Molecule type: DNA
 A:Residues: 1-166 <LAW>
 A:Cross-references: GB:J04835
 R:Hennesy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwein,
 J. Cell Biol. 108, 729-736, 1989
 A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in th
 A:Reference number: A30140; MUID:89139590
 A:Accession: A30140
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>
 A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
 Biochemistry 25, 8418-8425, 1986
 A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis c
 A:Reference number: A25812; MUID:87157592
 A:Accession: A25812
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-397 <KOB>
 A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354
 R:Dixit, V.M.; Hennesy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
 A:Reference number: A05172; MUID:86287276
 A:Accession: A05172
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>
 A:Cross-references: GB:M14336; NID:g340005; PIDN:AAA61237.1; PID:g53801
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
 J. Cell Biol. 118, 693-701, 1992
 A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin.
 A:Reference number: A42927; MUID:92348511
 A:Accession: A42927
 A:Molecule type: Protein
 A:Residues: 987-1003 <SUN>
 A:Note: Cys-992 is shown to have a free sulfhydryl

C:Genetics:
 A:Gene: GDB:THBS1; TSP1; TSP
 A:Cross-references: GDB:120438; OMIM:188060
 A:Map position: 19q15-15q15
 A:Introns: 23/1
 A:Note: the list of introns may be incomplete
 C:Complex: homotrimer, disulfide linked
 C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregat
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
 C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tr
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
 F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>
 F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF1>
 F:650-689/Domain: EGF homology <EGF2>
 F:926-928/Region: cell attachment (R-G-D) motif
 F:171-232/Disulfide bonds: #status predicted
 F:248, 360, 708, 1067/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:270, 274/Disulfide bonds: interchain #status predicted
 F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:1051/Binding site: carboxylate (Asn) (covalent) #status absent

Query Match 64.0%; Score 4228; DB 1; Length 1170;
 Best Local Similarity 61.6%; Pred. No. 4.5e-250;
 Matches 725; Conservative 170; Mismatches 267; Indels 14; Gaps 6;
 QY 1 MVWRL-VLLALWVWPSTQAGHQDKD-TTFDLFSINRKTIGAKQFRGDPGVPAYRFV 58
 Db 3 LAWGLGVFLMHVCTNRIPESGGDSNVDFELTGAARKSGRRVLKGPDPSSPAFRTE 62
 QY 59 RFDYPPVNNADLSKITIMRQKGEFFITAQKDGKSRGILLALEGGCLSGRQFEIVSN 118
 Db 63 DANLPPVPDDKFDQDLYDAVRTEKGFLLASLRQMKRTGTLALLERLDHSGQVESVYSN 122
 QY 119 GPADTDLITYWDGTRHVVVLEDVGLADSQMKNVTQVAGETYSLVHVCGLDLSFALDEP 178
 Db 123 GKAGTDLDSLIVQGHVVEALLATGQWKSTFLVQEDRAQLYIDCKENNELDVP 182
 QY 179 F---YEHLOAKESRMVAKSGARSHPRLGLQNLVHLPENSVEDILSKKQGGQGGAIN 235
 Db 183 IQSVFTRDLASIALRIAKGVND-NFQGLQNVRFVFGTTPEDILRNKCGSSSTVLLT 241
 QY 236 AISENTEFLRLGPHVTTTYVGPSSERRRERVCERSCCELGNMVQELSGHLVNVQPSNLK 295
 Db 242 L--DNVVVNGSSPAIRTYIHKTKDQLQAICGICDELSSMVLELRGLRTIVTTLQDSIR 299
 QY 296 RVSDNDFLWELIGPPKTRNNMSACWODGRFFAENETWVVDSTCTCTCKEFTCHQITC 355
 Db 300 KVTENKELANELRRPP-----LCYHNGVQYRNEETWVDSCTECHCONSVTICKKYSC 353
 QY 356 PPATCASPFVEGECPCSLHSVDGEGSWPAEWTCQSVTCGSGTOQRGSRCDVTSNVC 415
 Db 354 PIMPCSNATVPDGECCPCWPSDADGWSWSEWTSCSTSCGNGIQOGRGSCDSLNNRC 413
 QY 416 LPSIQTRACSLSKCDTRIRODGGWSHWSWSSCSVTCGVGNITRILCNSPVPMQGN 475
 Db 414 EGSSVQRTNCHIQCDDRFRKODGGWSHWSWSSCSVTCGVGNITRILCNLSPSPQMNKP 473
 QY 476 CKSGRETKACQAGPCPTDGRWSWSPWASCTVTTCAGIRTRVCNSPEPYQGGKACVG 535
 Db 474 CEGEARETKACKKDACPLNGMGWSPWPDICSVTCGGVQKRSLRCLNNPTQFGGKDCVG 533
 QY 536 DVQERQCMNRKSPVDGCLSNPCFPFQAQSSFPDGSWSCGCPVGLFNGTHGCHDDECA 595
 Db 534 DVTENQICNKQDCPIDGCLSNPCFAGVKCTSPYDGSWCKGACPPGYSGNGIQCTDVSCK 593
 QY 596 LVPDICEFSTSKVPRCVNTQPCFHLCPGPRYRGNQPVGVGLEAAKTEKQVCEPENPKDK 655

Db 594 EVPDACFNHNGEHCENTDPGYNCLPCPPRETGSQPPGQGVHATANKQVCKPRNPTDGS 653
Qy 656 THNCHKHAEICYLGHSDPMYKCECQTAGDGLIGESDLDGWNLNLCVATNATYHC 715
Db 654 THDCNNAKNCYLGHSDPMYKCECKPAGNAGIICGEDTLDGWPENLVCVANNATYHC 713
Qy 716 IKDNCPLNPSGQEDFDKDGIGDACDDDDNDGVTDEKDNCCOLLFNPROADYDKDEVGR 775
Db 714 KDCNCPNLSNGGEDIYKDGIGDACDDDDNDXIPDDRNCPPHYNPAQYDYDRDDVGR 773
Qy 776 CDCNCPYVHNPAQIDTNNEGDACSVDIDGDDVDNERNDCPPYVNTDQDRTDGDGVDHC 835
Db 774 CDCNCPYHNHPDQADTNNEGDACAADIDGDLNERNDCQYVYVNDQRTDMDGVGDQC 833
Qy 836 DANCPLVNPQDQVNDLVGDCQDNEDIDDDGHQNNQDNCPIISANQAADHRRDQGD 895
Db 834 DANCPLVNPQDQVNDLVGDCQDNEDIDDDGHQNNQDNCPIISANQAADHRRDQGD 893
Qy 896 CDPDDNDGVPDDRDNCRLVFNPDQEDLDGDRGDKDKDDFNDNDPDDDDVCPENNAIS 955
Db 894 CDHDDNDGIPDDKNCRLVNPDPQDKSDGDRGDKDKDDFNDNDPDDDDVCPENNAIS 953
Qy 956 ETDFRNFQVPLDPKQTTQIDPNMVIHQKELVQFANSDPGIAVGDFSGVDFSGTFY 1015
Db 954 ETDFRNFQVPLDPKQTTQIDPNMVIHQKELVQFANSDPGIAVGDFSGVDFSGTFY 1013
Qy 1016 VNTDRDDYAGFVFGYQSSRFYVWVKQVQTYWEDQTPRAYGSGVSLKVVNSTTGG 1075
Db 1014 INTERDDYAGFVFGYQSSRFYVWVKQVQTYWEDQTPRAYGSGVSLKVVNSTTGG 1073
Qy 1076 EHLRNALWHTGNTPGQVRLWHPDPRHGWKDYATYBWLHTRPKCTGYIRLVHESQVNA 1135
Db 1074 EHLRNALWHTGNTPGQVRLWHPDPRHGWKDYATYBWLHTRPKCTGYIRLVHESQVNA 1133
Qy 1136 DSGPIYDQYAGGRLGLFVFSQEMVYFSDLYKECRD 1171
Db 1134 DSGPIYDQYAGGRLGLFVFSQEMVYFSDLYKECRD 1169
RESULT 5
A40558
thrombospondin 1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence revision 05-Jun-1992 #text_change 20-Aug-1999
C:Accession: A40558; A37905; B42587; S68787
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:Title: Characterization of the murine thrombospondin gene.
A:Reference number: A40558; MUID:92128941
A:Accession: A40558
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <LAW>
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62465; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061
R:Bornstein, P.; Alf, D.; Devatavalu, S.; Framson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
A:Reference number: A37905; MUID:90375546
A:Accession: A37905
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <BOR>
A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA40431.1; PID:9554390
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.P.; Seidlin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683
A:Accession: B42587
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1152, 'p', 1154-1170 <LAH>
A:Cross-references: GB:M87276

A:Note: sequence extracted from NCBI backbone (NCBIP:81501)
R:Chen, H.; Aeschlimann, D.; Nowling, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin
A:Reference number: S68787; MUID:96234006
A:Accession: S68787
A:Molecule type: protein
A:Residues: 19-26, 'X', 28-37 <CHE>
C:Complex: homotrimer, disulfide linked
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology
F:1-18/Domain: signal sequence; glycoprotein; homotrimer
F:19-1170/Product: thrombospondin 1 #status predicted <SIG>
F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>
F:378-423/Domain: thrombospondin type 1 repeat homology <THR1>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:551-586/Domain: EGF homology <EGF>
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 63.4%; Score 4189.5; DB 2; Length 1170;
Best Local Similarity 61.8%; Pred. No. 1e-247;
Matches 723; Conservative 162; Mismatches 272; Indels 13; Gaps 5;

Qy 6 VLLALWVWSPQAGHDKDT-TDFLFSISNINRKTIGAKQFRGPDGVPAYRFRFDYIP 64
Db 9 VLLFLHMGCSNRPESGGDNGVDFIDELIGGARGRRLKVGQDLSFAFRLENANLIP 68
Qy 65 PYNADDLSKITIMROKEGFFLTAQLKQDGKSGRTLLALEGPGLSQRQFEIVSNGPADTL 124
Db 69 AVPDKKFQDLLDAVWADKGFIFLASLRQMKTRGTLLAVERDKNTGQIFSVSNGRAGTL 128
Qy 125 DLTWYWDGTRHYVSLDVGADLSQWKNVTVQVAGETYSYLVHGCGLDSDSFDALDEP---FYE 181
Db 129 DLSLSLPGKQVYVSEALLATGQWKSITLFVQEDRAQLYIDCKMESAEPLDPIOSIF 188
Qy 182 HLQAEKSRMYVAKGSARESHFGLLONVHLVFNSEVEDILSKKGCQGGCAEINASENT 241
Db 189 RLASVARLVRKAGDVND-NFGVLQNVRFVGTTFEDILRNKGCSSSTNVLTL--DNN 245
Qy 242 ETLRLGPHVTYVGPSPERRPERCERSEELGNMVOELSLHLVLPQSPNKRYSNDN 301
Db 246 VVNGSSPAIRTNVIGHKTKDLQALICGLSDELSSMWLELKLGLRTIVTLQDSIRKVTEN 305
Qy 302 QFLWELIGPPKTRNMSACWDGRFAENETWVVDSCCTCTCKKFKTICHOITCPCPATCA 361
Db 306 RELVSELKRPP-----LCFHNGVQYKNEEWTVDSCTECHCONSVTICKKVCSCPIMPCS 359
Qy 362 SPSEFVEGECPCSLHSDGEBGWSPAEWTCQSVTCGSGTQQRGRSCDVT-SNTCLGPSIQ 421
Db 360 NATVPDECCPRGWSDSADDDGWSPWSEWTSCSATNGIQGRSCDSLNNRCESSVQ 419
Qy 422 TRACLSKCDTRIRQDGGHSHSPSSCSVTGCVGNITRIRLNCSPVPQMGKNCKGSGR 481
Db 420 TRTCHIOECDKRFKQDGGHSHSPSSCSVTGCDGVITRIRLNCSPSPOMNGKPCGEAR 479
Qy 482 ETKACGACPCIDGRNPSHSPSACVTCAGGIRTRVNCNPEPQYGGKACVGDVQERQ 541
Db 480 ETKACKKDACPINGGWPSPMDICSVTCGGVQRRSLCNPPTQFGGKDCVGDVTENQ 539
Qy 542 MCKNRSCPVDGCLSNPCFPGAOCSPFDGSGSCGSPGVFLNGTHCEDLDECALVPDIC 601
Db 540 VCNKQDCPIDGCLSNPCFAGAKCTSPDGSWKACAPPGYSGNGICQCKVDDECKEVPDAC 599
Qy 602 FSTSKVPRCVNTQPGFCHLPCPPRYRGNPVGVLBAATEKQVCEPENPCDKTKINCHK 661
Db 600 ENHNGEHRCKNTDPGYNCLPCPPRFTGSPFGRGVEHAMANKQVCKPRNPCTDGTDCNK 659
Qy 662 HABCIVLGHFSDPMYKCECQTAGDGLIGCESDLDGWPENLVCVANNATYHCIDNCP 721
Db 660 NAKCNLYGHYSDPMYKCECKPGVAGNGIICGEDTLDGWPENLVCVANNATYHCKKDNCP 719

| Matches | 360; Conservative | 78; Mismatches | 183; Indels | 69; Gaps | 11; |
|--------------------------|--|--|-------------|----------|-----|
| Qy | 537 | VOERQMC---NKRSCVPDGLSLNPFPGAQCS---SPDGSWGSCGCPVGLNGTHCED | 530 | | |
| Db | 263 | IMEQVCGFHEQRS---HCSNPFPGVDMVEYEP--GYRCGPCPGQLQNGTHCSD | 316 | | |
| Qy | 591 | LDECALVPDTCFSTSKVPRCVNTQPGFHLPCPSPRTGNGOPVGVGLEAAKTEKQVCE--- | 647 | | |
| Db | 317 | INECAHA-DPCFPQSS---CINTMPGFHCEACPRGYKGTQVSGVDIYARASQVCNDID | 372 | | |
| Qy | 648 | -----PENPCKDKTHN-CHKHAEC | 665 | | |
| Db | 373 | ECNDGNGGCDPNISCTNTVGSFKGRCPLGFLGNSSQGLCPARTCHSPAHSCHLHAHC | 432 | | |
| Qy | 666 | IYLGHFSDPMYKCBCTGYAGDLGICEDSLDGWPNLNLVCAATNATYHCKDKNCPLPN | 725 | | |
| Db | 433 | LF---ERNGAVSCQNVGWAAGNVCGCTDIDGYDPOALPCHDN-NKHCQDKNCLLTPN | 488 | | |
| Qy | 726 | SGQEDFKDIGDACDDDDNDGVYDDEKNCQLLFPNRQADYDKDEVGRDCNCPVYHNP | 785 | | |
| Db | 489 | SGQEDADNDGVGQCDQDDADGSDGINKVEDNCRFLPNKDKQNSDTSFGDACDNCNPVNN | 548 | | |
| Qy | 786 | AQIDTANNGEDACSDIDGDDYFNERDNCPPYNTDRTDGDGVGDHCDNCPLVHNP | 845 | | |
| Db | 549 | DQKDTGNGEGDADNDVDGSGIPNLGDNCKPVPNLQTDREDDGVGDACDSCPEMSNPT | 608 | | |
| Qy | 846 | QTDVDNLVGDQCDNNEDIDDGHONNDNCPIYSNANQADHDDGOGDACDPDDNDGV | 905 | | |
| Db | 609 | QTDADSLVGDVCDTNEGSDGHDQDKNCQPOLPNSSQLDSNDGLGDECDGDDNDGI | 668 | | |
| Qy | 906 | PD-----DRDNCRLVFNPDQEDLDGDRGDIKCDKDDFNNDIPDIDDDYCPENNAIS | 961 | | |
| Db | 669 | PDYVPPGPDNCRVLPNPNQKSDGNGVGDVCEDDFNDADVVDPLDVCESAETLTDFA | 728 | | |
| Qy | 962 | FQWYPLDPKCTQIDPNWVLRHOGKELVQATNSDPGTATGVDFGSDVSGFYVNTDRD | 1021 | | |
| Db | 729 | YQTWLLDPEGDAQIDPNWVVLNMGMEIVQPMNSDPLGAVGYTAFNGVDEGTFHNTVTD | 788 | | |
| Qy | 1022 | DDYAGFYFGQSRSREVVVWVKQVTTQTYWBDQPTRAYGSGVSLKVVNSTGTGHELRNA | 1081 | | |
| Db | 789 | DDYAGFLFSYQDSGREVVVWVKQTEQTYWQATPFRAVAPGLQKAVTSVSGPGEHRNA | 848 | | |
| Qy | 1082 | LWHTGNTPGVRTLWHDPRNIGWKDYATRWHLTHRPKGYIRVLVHEGKQVMADSGPY | 1141 | | |
| Db | 849 | LWHTGHTPDQVRLTLWTDPRNVGRWDRKTSYRWQLLHRRPQGYIRVKLYEGPQLVADSGVII | 908 | | |
| Qy | 1142 | DQYVAGGRGLGVFSCQEMVYFSLKYECD | 1171 | | |
| Db | 909 | DTSMRGRGLGVFCFQENILWNLQYKCD | 938 | | |
| RESULT | 8 | | | | |
| A46016 | | | | | |
| thrombospondin 3 - mouse | | | | | |
| C:Species: | Mus musculus (house mouse) | | | | |
| C:Date: | 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000 | | | | |
| C:Accession: | A46016; A44124; I55398; S68788; S72433 | | | | |
| R:Bornstein, | P.; Devarayalu, S.; Edelhoff, S.; Disteche, C.M. | | | | |
| Genomics | 15, 607-613, 1993 | | | | |
| A:title: | Isolation and characterization of the mouse thrombospondin 3 (Thbs3) gene. | | | | |
| A:Reference | number: A46016; MUID:93224149 | | | | |
| A:Accession: | A46016 | | | | |
| A>Status: | preliminary | | | | |
| A:Molecule | type: mRNA | | | | |
| A:Residues: | 1956 <BOR> | | | | |
| A:Cross-references: | GB:I04302; NID:g202200; PIDN:AAA0497.1; PID:g202201 | | | | |
| A:Note: | sequence extracted from NCBI backbone (NCBIN:129415; NCBI:P.129416) | | | | |
| R:Vos, | H.L.; Devarayalu, S.; de Vries, Y.; Bornstein, P. | | | | |
| J. Biol. Chem. | 267, 12192-12196, 1992 | | | | |
| A:title: | Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene family. | | | | |
| A:Reference | number: A44124; MUID:92291102 | | | | |
| A:Accession: | A44124 | | | | |
| A>Status: | preliminary | | | | |

A;Molecule type: DNA
A;Residues: 517-956 <VOS>
A;Cross-references: GB:M86620; NID:g201987; PIDN:AAA0430.1; PID:g201989
A;Experimental source: BALB/c
A;Note: sequence extracted from NCBI backbone (NCBIN:106634, NCBIN:106636, NCBIN:106106)
)
A;Qabar, A.N.; Lin, Z.; Wolf, F.W.; O'Shea, K.S.; Lawler, J.; Dixit, V.M.
J. Biol. Chem. 269, 1262-1269, 1994
A;Title: Thrombospondin 3 is a developmentally regulated heparin binding protein.
A;Reference number: I55398; MUID:94117438
A;Accession: I55398
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-432; 'F', 434-562, 'D', 564-719, 'E', 721-870, 'W', 872-956 <RES>
A;Cross-references: GB:I24434; NID:g402718; PIDN:AAA04033.1; PID:g402719
R;Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 35-41, 1996
A;Title: Expression and initial characterization of recombinant mouse thrombospondin
A;Reference number: S68787; MUID:96234006
A;Accession: S68788
A;Molecule type: protein
A;Residues: 'X', 23-28, 'X', 30, 'X', 32 <CHE>
R;Collins, M.; Bornstein, P.
Nucleic Acids Res. 24, 3661-3669, 1996
A;Title: SP1-binding elements, within the common metaxin-thrombospondin 3 intergenic
A;Reference number: S72433; MUID:97025352
A;Accession: S72433
A;Molecule type: DNA
A;Residues: 1-26 <COL>
A;Cross-references: EMBL:U66257; NID:g1575552; PIDN:AAC52819.1; PID:g1575554
C;Genetics:
A;Gene: Thbs-3
C;Complex: homopentamer, disulfide linked
C;Superfamily: thrombospondin 3; EGF homology
C;Keywords: calcium binding; glycoprotein; homopentamer
F:374-412/Domain: EGF homology <EGF>
F:310,407,644,937/Binding site: carbohydrate (Asn) (covalent) #status predicted

[illegible]

Db 669 PDYIPPGDNCRLVVPNPNDKSDGNGVGDVCEDDFDNDVAVDPDLVCPESAGVTLTDFRA 728

Qy 962 FQWYPLDPKGTQIDFNWVIRHOGKELVOTANSDFGIAVGEFGSDVDFSGTFVYNTDRD 1021

Db 729 QYVILDPEDQAIDFNWVVLNAGOEIVQTMNSDGLAVGTAFNGVDVEGTFHFVNTVD 788

Qy 1022 DDYAGVFVGQSSRRFYVVMKQVQTYWEDOPTRAYGYSGVSLKVNSTTGTGHELRNA 1081

Db 789 DDYAGFLSYQDSGRFYVVMKQTEQTYWQATPFRAVAQPGQLKAVTISISGPEHLRNA 848

Qy 1082 LWHGTNTPGQVRYLWHDPRNIGKDYATYRWHLTHRPKTYIRVLVHEGKQVMADSGPIY 1141

Db 849 LWHGHTPDQVRLWTDPRNVGLRDKTYSRWLLRHPQVGYIRKLYEGPOLVADSGVII 908

Qy 1142 DQYAGGRLGLVFQSEWYFSDLKYECDR 1171

Db 909 DTSMRGRLGVCFQSOENIWSNLQYRCND 938

RESULT 9

A45441

thrombospondin 4 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000

C:Accession: A45441

R:Lawler, J.; Duquette, M.; Whittaker, C.A.; Adams, J.C.; McHenry, K.; Desin

J. Cell Biol. 120, 1059-1067, 1993

A:Title: Identification and characterization of thrombospondin-4, a new mem

A:Reference number: A45441; MUID:93163109

A:Accession: A45441

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-955 <LAW>

A:Cross-references: GB:219091; NID:g288777; PIDN:CAA79518.1; PID:g288778

A:Note: sequence extracted from NCBI backbone (NCBIN:124858, NCBIPI:124860)

C:Superfamily: thrombospondin 3; EGF homology

F:285-319/Domain: EGF homology <EGF>

Query Match 28.8%; Score 1899; DB 2: Length 955;

Best Local Similarity 51.9%; Pred. No. 4.6e-108;

Matches 348; Conservative 79; Mismatches 184; Indels 60; Gaps

Qy 549 PVDGCLSNCPFGAQCSPFDGSSWCGSCPVGFLNGTHCEDLDECALPDICFSTSKVP 608

Db 281 PKPRCDATSCFRGVRCID-TEGFGCGCPCEYTGIVDFVDECRINP-CFLG---V 334

Qy 609 RCVNTQPGFHCLPCPPRYRGNDPVGLEAKTEKV----- 645

Db 335 RCINTSPGFKCESCPPGYTGSTIQIGINFARQNKQVCTDTNECENGRNGGCTSNSLCIN 394

Qy 646 -----CEPENPKDKTHCHKHAEICYLGHFSDPMYKCEOTGY 684

Db 395 TMGSFRCGGCKPGYGVGDQTKGCKPKSRHGNPCHASAQC---SEKGDGVTCVSGV 451

Qy 685 AGDLICIGEDSOLDGPNLNLVCATNATYHCKNCPHLPNSGQEDFKDGIQDADDDD 744

Db 452 AGNYLCGRDIDIDGYPDALCPDK---NCKKNCVYVPSNGGEDTDKNIGDCADEDA 508

Qy 745 DNDGVTDEKDNQCLLFNPRQADYDKVEYDRCDCNCPYVHNPAQIDTNNNGEGSACSVID 804

Db 509 DGDGILLNEQDNCVLAANI DOKNSDQIFGDACDNCRLFLNNDQRTDNDKGDACDDMD 568

Qy 805 GDDVFERDNCPPYVNTDQRTDGDGVGDHCDNCPLVHNPQDTVDNDLVGDQCDNNEDI 864

Db 569 GDGIKILINDCQRPVNDQKRDGSDGVIDCSDCPDIINPNQSIDNDLVGSDCTNQDS 628

Qy 865 DDGHHONNDNCPIYSNANOADHRDGDGDACDPPDDNDGVPD---DRNCRCLVFNPDQ 920

Db 629 DGDGHQSDTNCPTVINSNQLDITDKDGDGECDDDDNDGIPDTPVPGPNCKNLVFPNGQ 688

Qy 921 EDLDGSDGRGICKDDFDNDNIPDIDVCPENNAISETDFRNFQMVPLDPKGTQTIDPNWY 980

Db 689 EDDNNNDGVDYCEADFQDQTDVIDRIDVCPENAEITITLDFRAYQTIVLVPDQEGDAQIDPNWI 748

Qy 981 IRHOGKELVOTANSDPGIAVGFDEFSGVDFTGTFYVNTDRDDDYAGFVFGYQSSSRFYV 1040

Db 749 VLNQGMELVQTMNSDPLGAVGYAFNGVDFEGTFHVTMTDDDYAGFVFGYQSSSFYV 808

Qy 1041 MWKQVOTYWEODPTRAIVGYSGVSLKVNVTGTGTEHLRNALNHHGTNPFGVQVRLTWHDP 1100

Db 809 MWKQTEQTYQATPFRAVAPGPIQLKAVAKSGSPGHEHLRNALNHHGTNDQVRLNWKDPR 868

Qy 1101 NIGMKDVTAYRWHLTHRPKTYIRVLVHBEKQVMAADSGPIYDOTYAGGRLGLGVFVSQEMV 1160

Db 869 NVGMKDKVSTRNFWLQHRPQVGYIRARFYEGTELVDADSGVTDTTMRGRLGVFCFSQENI 928

Qy 1161 YFSDLKYECRD 1171

Db 929 IWSNLKYRCND 939

RESULT 10

TSHUP4

thrombospondin 4 precursor - human

C:Species: Homo sapiens (man)

C:Date: 06-Jan-1995 #sequence_revision 11-Aug-1995 #text_change 13-Aug-1999

C:Accession: A55710; S36089

R:Lawler, J.; Mchenry, K.; Duquette, M.; Derick, L.

J. Biol. Chem. 270, 2809-2814, 1995

A:Title: Characterization of human thrombospondin-4.

A:Reference number: A55710; MUID:95155352

A:Accession: A55710

A:Molecule type: mRNA

A:Residues: 1-961 <LAW>

A:Cross-references: EMBL:Z19585; NID:g311625; PIDN:CAA79635.1; PID:g311626

A:Note: authors translated the codon GTG for residue 616 as Ser

C:Genetics:

A:Gene: GDB:THBS4

A:Cross-references: GDB:463011; OMIM:600715

A:Map position: 1q21-1q23

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet

C:Superfamily: thrombospondin 3; EGF homology

C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium bind

E:1-21/Domain: signal sequence #status predicted <SIG>

F:22-961/Product: thrombospondin 4 #status predicted <MAT>

F:290-324/Domain: EGF homology <EGF>

F:330-362/Domain: EGF homology <EGF>

F:562-564/Region: cell attachment (R-G-D) motif

F:303/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predic

F:343/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:612,941/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 28.6%; Score 1890; DB 1; Length 961;

Best Local Similarity 53.1%; Pred. No. 1.6e-107;

Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps

Qy 549 PVDGCLSNPCFPGAQGSFPDGSWCGSCPVLGNGTHCEDLDECALVPDICTSTSKVP 608

Db 286 PPRCDSPNCFRGVQCTDSRDG-FQCGPCEGTNGTICIDVDECKYHP--CYPG---V 339

Qy 609 RCVNTQPGFHCILCPPRYRGNQPVGVGLAAKTEKQVC-----EPEN----- 650

Db 340 HCINLSPGFRCDACPVGTGPMVQGVGSFSAKSNKQVCTDIDECRNGACVPNSICVNTLG 399

Qy 651 -----PCK-----DKTNCHNRHABCIYLGHSDFPMWKECQGTGAG 686

Db 400 SYRCGPCKPGYTDQIRGCKVERNCNPNELNPCSVNAQCI--EERQGDVTCVGVGWAG 456

Qy 687 DGLICGEDSLDQGNWNLNLCVATNATYHCIKDNCPLNSQGEFDDKDGIGDADDDDDN 746

Db 457 DGYTCRKDVDDISYDEEELPCFSAR---NCKKDNCKYVPSNGQEDADRIGDGEDADG 513

| | | | | | |
|----|-----|--|----------------------------------|------|-----|
| QY | 99 | TLLALEGPGLSQR----- | -----QFEIVSNGPADTLDITYWIDGT----- | RHVS | 138 |
| | | | | | |
| | | | | | |
| | | | | | |
| Db | 897 | TCSATCGEGLOSRSCRRGSGCTEDDASQTRRCVNGPCEHSYLT-WSEWTTCTCSCSFD | | | 955 |

F;408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Db 295 EGGGCF-----GVLEEGGROCNRACGPGRTSSRSOSLRSTDDARRRREYGDLEIQQ-- 344

QY 282 GLHVLVNPSENKRVSNQFLWELIGGP-PKTRNMSACWDGRFFAENETWVVDSCTT 340
Db 345 -----FGFPAPQTGDPAA-----EWSPMWCVCSSTC 370
QY 341 -----CTCKKFTICHQITCPATCASPFSVEGECPCSLHSDVDGEGWSPWAEW 390
Db 371 GEGWQTRTRCVSSYST-----QCSGPLREORLCNNSAVCPVHG--ANDEWSPW 418
QY 391 TQCSVTCSGTQORGRSC---DVTNLTCLGPSIQTRACSLSKCDTRIRQDGGWHSWSPWS 447
Db 419 SLASSTCGRFRDTRTCRPPQFGGNPCGEPKQTKFCNIALCPGR-AVDGNWNNWSSWS 477
QY 448 SCSVTCGVGNITRIRLNSVPQMGKNCKGSGRETRACOGAPCPIDGRHSPWSPHSACT 507
Db 478 ACSASCQSGRQRTRECNG--PSYGGAECCGHHVYETRDCLQOCPPVDGKQWASWGS 535
QY 508 VTCAGGIRERTRVCNSPEPOYGGKACVGYOERQMCNKRSCP--DGCLSNPCFPGGAQCS 565
Db 536 VTCAGSORRERVCSG--PFFGGAACOGPODEYRQCGTORCPEPHEIC-DEDNFGAVIWK 592
QY 566 SFPDGSWSCGSCVPVFLNGT-----HCEDLDE 593
Db 593 ETPAGEVAAVRCP-----RNATGLILRRCE-LDE 620

Search completed: August 9, 2002, 10:02:15
Job time: 171 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:01:40 ; Search time 15.18 seconds
(without alignments)
2989.415 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVLLALWVPSQAGH.....FVFSQEMVYFSDLKYECRDI 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 6570 | 99.5 | 1172 | 1 TSP2_HUMAN | P35442 homo sapien |
| 2 | 6068 | 91.9 | 1170 | 1 TSP2_BOVIN | Q95116 bos taurus |
| 3 | 5920 | 89.6 | 1172 | 1 TSP2_MOUSE | Q03350 mus musculus |
| 4 | 5481 | 83.0 | 1178 | 1 TSP2_CHICK | P35440 gallus gall |
| 5 | 4228 | 64.0 | 1170 | 1 TSP1_HUMAN | P07996 homo sapien |
| 6 | 4228 | 63.9 | 1170 | 1 TSP1_BOVIN | Q28178 bos taurus |
| 7 | 4189.5 | 63.4 | 1170 | 1 TSP1_MOUSE | P35441 mus musculus |
| 8 | 4189 | 63.4 | 1173 | 1 TSP1_XENLA | P35448 xenopus lae |
| 9 | 1993 | 30.2 | 757 | 1 COMP_HUMAN | P49747 homo sapien |
| 10 | 1968 | 29.8 | 755 | 1 COMP_RAT | P35444 rattus norv |
| 11 | 1946.5 | 29.5 | 956 | 1 TSP3_HUMAN | P49746 homo sapien |
| 12 | 1936.5 | 29.3 | 956 | 1 TSP3_MOUSE | Q05895 mus musculus |
| 13 | 1899 | 28.8 | 955 | 1 TSP4_XENLA | Q06441 xenopus lae |
| 14 | 1890 | 28.6 | 961 | 1 TSP4_HUMAN | P35443 homo sapien |
| 15 | 1881 | 28.5 | 980 | 1 TSP4_RAT | P49744 rattus norv |
| 16 | 1480 | 22.4 | 439 | 1 COMP_BOVIN | P35445 bos taurus |
| 17 | 439.5 | 6.7 | 437 | 1 PROP_MOUSE | P11680 mus musculus |
| 18 | 424 | 6.4 | 1384 | 1 BAI1_HUMAN | O14514 homo sapien |
| 19 | 418.5 | 6.3 | 810 | 1 BAI1_RAT | O62919 rattus norv |
| 20 | 418 | 6.3 | 1522 | 1 BAI3_HUMAN | O60242 homo sapien |
| 21 | 416 | 6.3 | 469 | 1 PROP_HUMAN | P27918 homo sapien |
| 22 | 415 | 6.3 | 1074 | 1 SMSA_HUMAN | Q13591 homo sapien |
| 23 | 407.5 | 6.2 | 470 | 1 PROP_CAVPO | Q64181 cavia porce |
| 24 | 391 | 5.9 | 1077 | 1 SMSA_MOUSE | O62217 mus musculus |
| 25 | 390 | 5.9 | 1093 | 1 SMSB_MOUSE | Q60519 mus musculus |
| 26 | 389.5 | 5.9 | 816 | 1 NEL2_MOUSE | Q61220 mus musculus |
| 27 | 388.5 | 5.9 | 810 | 1 NEL1_HUMAN | Q92832 homo sapien |
| 28 | 384.5 | 5.8 | 1572 | 1 BAI2_HUMAN | O60241 homo sapien |
| 29 | 379.5 | 5.7 | 816 | 1 NEL2_RAT | Q62918 rattus norv |
| 30 | 366.5 | 5.5 | 816 | 1 NEL_CHICK | Q90827 gallus gall |
| 31 | 361.5 | 5.5 | 816 | 1 NEL2_HUMAN | Q99435 homo sapien |
| 32 | 343.5 | 5.2 | 2531 | 1 NTC1_MOUSE | Q01705 mus musculus |
| 33 | 340.5 | 5.2 | 2531 | 1 NTC1_RAT | Q07008 rattus norv |

| | | | | | |
|----|-------|-----|------|--------------|---------------------|
| 34 | 329.5 | 5.0 | 807 | 1 FSPQ_RAT | P35446 rattus norv |
| 35 | 326.5 | 4.9 | 2524 | 1 NOTC_XENLA | P21783 xenopus lae |
| 36 | 308 | 4.7 | 803 | 1 FSPQ_XENLA | P35447 xenopus lae |
| 37 | 307 | 4.6 | 2871 | 1 FBNI_MOUSE | Q61534 mus musculus |
| 38 | 306.5 | 4.6 | 2444 | 1 NTC1_HUMAN | P46531 homo sapien |
| 39 | 305.5 | 4.6 | 2703 | 1 NOTC_DROME | P07207 drosophila |
| 40 | 300.5 | 4.5 | 2437 | 1 NOTC_BRARE | P46530 brachydanio |
| 41 | 300 | 4.5 | 2871 | 1 FBNI_HUMAN | P35555 homo sapien |
| 42 | 297 | 4.5 | 2871 | 1 FBNI_BOVIN | P98133 bos taurus |
| 43 | 296 | 4.5 | 2139 | 1 CRB_DROME | P10040 drosophila |
| 44 | 293.5 | 4.4 | 2871 | 1 FBNI_PIG | Q9TV36 sus scrofa |
| 45 | 292 | 4.4 | 1629 | 1 ATS9_HUMAN | Q9P2N4 homo sapien |

ALIGNMENTS

| RESULT | 1 |
|------------|--|
| TSP2_HUMAN | |
| ID | TSP2_HUMAN |
| AC | P35442; STANDARD; PRT; 1172 AA. |
| DT | 01-JUN-1994 (Rel. 29, Created) |
| DT | 01-JUN-1994 (Rel. 29, Last sequence update) |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) |
| DE | Thrombospondin 2 precursor. |
| GN | THBS2 OR TSP2. |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| OX | NCBI_TaxID=9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=94010892; PubMed=8406456; |
| RA | Labell T.L., Byers P.H.; |
| RT | "Sequence and characterization of the complete human thrombospondin 2 |
| RT | cDNA: potential regulatory role for the 3' untranslated region."; |
| RL | Genomics 17:225-229(1993). |
| RN | [2] |
| RP | SEQUENCE OF 560-1172 FROM N.A. |
| RC | TISSUE= Fibroblast; |
| RX | MEDLINE=92217961; PubMed=1559694; |
| RA | Labell T.L., McGookey Milewicz D.J., Distche C.M., Byers P.H.; |
| RT | "Thrombospondin II: partial cDNA sequence, chromosome location, and |
| RT | expression of a second member of the thrombospondin gene family in |
| RT | humans."; |
| RL | Genomics 12:421-429(1992). |
| CC | -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND |
| CC | CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, |
| CC | LAMININ AND TYPE V COLLAGEN. |
| CC | -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED. |
| CC | -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY. |
| CC | -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN. |
| CC | -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS. |
| CC | -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS. |
| CC | -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ |
| CC | or send an email to license@isb-sib.ch). |
| CC | ----- |
| CC | EMBL; L12350; AAA03703.1; - |
| DR | EMBL; M81339; -; NOT_ANNOTATED_CDS. |
| DR | PIR; A42173; A42173. |
| DR | HSSP; P35555; 1EMN. |
| DR | MIM; 188061; - |
| DR | InterPro; IPR000561; EGF-like. |
| DR | InterPro; IPR001881; EGF_Ca. |
| DR | InterPro; IPR001791; Laminin_G. |
| DR | InterPro; IPR000884; TSPI. |

```
DR InterPro: IPR003129; TSPN:
DR InterPro: IPR001007; VWFC:
DR InterPro: IPR003367; tsp_3:
DR Pfam: PF00008; EGF_2:
DR Pfam: PF02210; TSPN; 1:
DR Pfam: PF00090; tsp_1; 3:
DR Pfam: PF02412; tsp_3; 9:
DR Pfam: PF00093; VWC; 1:
DR SMART: SM00181; EGF; 3:
DR SMART: SM00209; TSP1; 3:
DR SMART: SM00210; TSPN; 1:
DR SMART: SM00214; VWC; 1:
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1:
DR PROSITE: PS00092; TSP1; 3:
DR PROSITE: PS01208; VWC; 1:
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 928 930
FT DISULFID 286 286
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
FT DISULFID 652 665
FT DISULFID 659 678
FT DISULFID 680 691
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Query Match 99.5%; Score 6570; DB 1; Length 1172;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MYRLLVLLALWYWPSTQAGHQDKDTTDLFSISINRKRTIGAKQFRGPDGVPAYRFRVF 60
DB 1 MYRLLVLLALWYWPSTQAGHQDKDTTDLFSISINRKRTIGAKQFRGPDGVPAYRFRVF 60
QY 61 DYIPPNADDLSKITKIMRQKEGFFLTALQKODGKSRGTLTALLEGPGLSQRQFEIVSNGP 120
DB 61 DYIPPNADDLSKITKIMRQKEGFFLTALQKODGKSRGTLTALLEGPGLSQRQFEIVSNGP 120
QY 121 ADTLDTTWIDGTRHVSLEDEGLADSQWKNVTVQVAGETYSLSHVGCIDLIDSFALDEPFY 180
DB 121 ADTLDTTWIDGTRHVSLEDEGLADSQWKNVTVQVAGETYSLSHVGCIDLIGPVALDEPFY 180
```

RESULT 2

TSP2_BOVIN

ID TSP2_BOVIN

STANDARD;

PRT; 1170 AA.

AC O95116; Q28180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
 DE (CISP).
 GN THS2 OR TSP2 OR TSP-2.
 GN Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID:9913;
 [1]
 RN SEQUENCE FROM N.A.
 RA Danik M., Chinn A., Lafuillade M., Keramidas M., Aquesse-Germon S.,
 RA Lafuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
 RA Feige J.J.;
 RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
 RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
 RT hormone in adrenocortical cells";
 RL J. Cell. Physiol. 167:164-172(1996).
 RN [3]
 RP SEQUENCE OF 1-522 FROM N.A.
 RX MEDLINE-96331130; PubMed-8698834;
 RA Lafuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
 RA Feige J.J.;
 RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
 RT TGF-beta";
 RL Tissue-Aortic endothelium;
 RC Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
 RA "Cloning and sequencing of bovine thrombospondin stimulatory effect of
 RT TGF-beta";
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOPOLYMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X96540; CAA65385.1; -;
 DR EMBL; X87620; CAA60952.1; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR003129; TSPN.
 DR InterPro: IPR001007; VWFC.
 DR InterPro: IPR003367; tsp_3.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF02412; tsp_3; 9.
 DR Pfam; PF00093; vwc; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18 POTENTIAL..
 FT CHAIN 19 1170 THROMBOSPONDIN 2.
 FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 318 375 VWFC.
 FT DOMAIN 379 430 TSP TYPE-1 1.
 FT DOMAIN 435 491 TSP TYPE-1 2.
 FT DOMAIN 492 546 TSP TYPE-1 3.
 FT DOMAIN 547 587 EGF-LIKE 1.
 FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 646 690 EGF-LIKE 3.
 FT DOMAIN 723 758 TSP TYPE-3 1.
 FT DOMAIN 759 781 TSP TYPE-3 2.
 FT DOMAIN 782 817 TSP TYPE-3 3.
 FT DOMAIN 818 840 TSP TYPE-3 4.
 FT DOMAIN 841 878 TSP TYPE-3 5.
 FT DOMAIN 879 914 TSP TYPE-3 6.
 FT DOMAIN 915 950 TSP TYPE-3 7.
 FT DOMAIN 951 1170 C-TERMINAL.
 FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).
 FT DISULFID 551 562 BY SIMILARITY.
 FT DISULFID 556 572 BY SIMILARITY.
 FT DISULFID 575 586 BY SIMILARITY.
 FT DISULFID 592 608 BY SIMILARITY.
 FT DISULFID 599 617 BY SIMILARITY.
 FT DISULFID 620 644 BY SIMILARITY.
 FT DISULFID 650 663 BY SIMILARITY.
 FT DISULFID 657 676 BY SIMILARITY.
 FT DISULFID 678 689 BY SIMILARITY.
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 535 535 A -> V (IN REF. 3).
 FT CONFLICT 748 748 S -> T (IN REF. 3).
 SQ SEQUENCE 1170 AA; 129862 MW; 9CF1BF55B89A051 CRC64;

Query Match 91.9%; Score 6068; DB 1; Length 1170;
 Best Local Similarity 90.0%; Pred. No. 0;
 Matches 1055; Conservative 62; Mismatches 53; Indels 2; Gaps 1;

QY 1 MWRLVLLALWVWPSTQAGHQDKDTTDFLSISINIKRTIGAKQFRGDPGVPAYRVRF 60
 DB 1 MLWPLLLALWVWPSTQAGHQDKDTTDFLSISINIKRTIGAKQFRGDPGVPAYRVRF 60
 QY 61 DYIPPNADLSKITKIMROKGFLLTAQLKQDGKSGRTLLALRGLSGLSRQFEIVSNGP 120
 DB 61 DYIPPNADLSKITKIMROKGFLLTAQLKQDGKSGRTLLALRGLSGLSRQFEIVSNGP 120
 QY 121 ADTLDLTWIDGTRHVVSLVDVGLADSQWKNVTVQVAGETYSLVHVGCDLDSFALDEPFF 180
 DB 121 ADTLDLTWIDGTRHVVSLVDVGLADSQWKNVTVQVAGETYSLVHVGCDLDSFALDEPFF 180
 QY 181 EHLQAEKSRMYVAKGSAHSHFRGLQNVHLVPENSVEDILSKGCGOOGAEINAISEN 240
 DB 181 EHLQAEKSRMYVAKGSAHSHFRGLQNVHLVPENSVEDILSKGCGOOGAEINAISEN 240
 QY 241 TETLRGLPHVTTVEYVGPSSRRRPEVCERSCCEELGNMVGQELSLHVLNQPSENLKRYSND 300
 DB 241 TETLRGLPHVTTVEYVGPSSRRRPEVCERSCCEELGNMVGQELSLHVLNQPSENLKRYSND 300
 QY 301 NQFLWELIGGPPKTRNNSACWQGRFFAENETWVVDSCCTCTCKKFKTICHQITCPTATC 360
 DB 301 NQFLWELIGGPPKTRNNSACWQGRFFAENETWVVDSCCTCTCKKFKTICHQITCPTATC 360

| | | | | | |
|----------------------------|----------|---|--|-------------------------------------|---------------------------------------|
| FT | DOMAIN | 843 | 880 | TSP TYPE-3 5. | |
| FT | DOMAIN | 881 | 916 | TSP TYPE-3 6. | |
| FT | DOMAIN | 917 | 952 | TSP TYPE-3 7. | |
| FT | DOMAIN | 953 | 1172 | C-TERMINAL. | |
| FT | SITE | 928 | 930 | CELL ATTACHMENT SITE (POTENTIAL). | |
| FT | DISULFID | 266 | 266 | INTERCHAIN (PROBABLE). | |
| FT | DISULFID | 270 | 270 | INTERCHAIN (PROBABLE). | |
| FT | DISULFID | 553 | 564 | BY SIMILARITY. | |
| FT | DISULFID | 558 | 574 | BY SIMILARITY. | |
| FT | DISULFID | 577 | 588 | BY SIMILARITY. | |
| FT | DISULFID | 594 | 610 | BY SIMILARITY. | |
| FT | DISULFID | 601 | 619 | BY SIMILARITY. | |
| FT | DISULFID | 622 | 646 | BY SIMILARITY. | |
| FT | DISULFID | 652 | 665 | BY SIMILARITY. | |
| FT | DISULFID | 659 | 678 | BY SIMILARITY. | |
| FT | DISULFID | 680 | 691 | BY SIMILARITY. | |
| FT | CARBOHYD | 151 | 151 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 316 | 316 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 330 | 330 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 457 | 457 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 584 | 584 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 710 | 710 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| FT | CARBOHYD | 1069 | 1069 | N-LINKED (GLCNAC. . .) (POTENTIAL). | |
| SQ | SEQUENCE | 1172 | AA; 129911 MW; 7CE8E4E8599822AB CRC64; | | |
| Query Match | | | | | 89.6%; Score 5920; DB 1; Length 1172; |
| Best Local Similarity | | | | | 88.5%; Pred. No. 0; |
| Matches 1036; Conservative | | | | | 57; Mismatches 78; Indels 0; Gaps 0; |
| QY | 1 | MVRLVLLALWVWPSTQAGHDKDTTFDLFSINIRKTIKAGKQFRGPDGPAYRFRF | 60 | | |
| DB | 1 | MLWALLALGIGPRASAGDHVKDTSFSLFSINIRKTIKAGKQFRGPDGPAYRFRF | 60 | | |
| QY | 61 | DYIPPNADLLSKITIMRKEGFFTAQLKQDKSRGTLALLEGPLSOROFEIYVNSGP | 120 | | |
| DB | 61 | DYIPPNADLLSKITIMRKEGFFTAQLKQDKSRGTLALLEGPLSOROFEIYVNSGP | 120 | | |
| QY | 121 | ADTLDTYWDGRRHVSLEVDGLADSQMKNVTQVAGETYSLVHVGCDLIDSPALDEPY | 180 | | |
| DB | 121 | GDTLDLNLYWVEGQHTNFLEVDGLADSQMKNVTQVADSTYSLVHVGCDLIDSTLEPEY | 180 | | |
| QY | 181 | EHLOAKSRMYAVKASRSHFGLLQNVHLVFPENVEDILSKKQCGQGAENAISEN | 240 | | |
| DB | 181 | EQLEVDKSRMYAVKASRSHFGLLQNVHLVFPENVEDILSKKQCGQGAENAISEN | 240 | | |
| QY | 241 | TETLRGLPHVTTEYVGSSEPRPEVCESEELGNVQVLSGLVHLVNPQSENLRKVSND | 300 | | |
| DB | 241 | TETLRGLPHVTTEYVGSSEPRPEVCESEELGNVQVLSGLVHLVNPQSENLRKVSND | 300 | | |
| QY | 301 | NQFLWELIGGPPKTRMSACWDGRFAENETWVDSCTCTCKKFKTCHQITCPTATC | 360 | | |
| DB | 301 | NOFLWELIGGPPKTRMSACWDGRFAENETWVDSCTCTCKKFKTCHQITCPTATC | 360 | | |
| QY | 361 | ASPSFVEGECPCSLHSVDEGWSWAEWTCQSVTCGSGTOQRGSCDVTSNTCLGPSI | 420 | | |
| DB | 361 | ANPSFVEGECPCSLHSVDEGWSWAEWTCQSVTCGSGTOQRGSCDVTSNTCLGPSI | 420 | | |
| QY | 421 | QTRACSLSKCDTRIRDDGGHSHWSPSSCVTCGVGNITRILCNLSPVPMQGNCKKGS | 480 | | |
| DB | 421 | QTRACSLSKCDTRIRDDGGHSHWSPSSCVTCGVGNITRILCNLSPVPMQGNCKKGS | 480 | | |
| QY | 481 | RETKACQAPCPIDGRWSPWSPSACTVTCAGGIRTRVCNSPEPYGGKACGVQDER | 540 | | |
| DB | 481 | RETKACQAPCPIDGRWSPWSPSACTVTCAGGIRTRVCNSPEPYGGKACGVQDER | 540 | | |
| QY | 541 | QMCNKRKCPVDCGLNCPGPAOCSSFPDGSWSCGCPVGLNGHTHCEDLDBCAVPTDI | 600 | | |
| DB | 541 | QMCNKRKCPVDCGLNCPGPAOCSSFPDGSWSCGCPVGLNGHTHCEDLDBCAVPTDI | 600 | | |
| QY | 601 | CFSTKVPKRCVNTQPGFCHLCPKPPRYRGNOQVGVGLEAAKTRQVCEPENCKDKTHNCH | 660 | | |
| DB | 601 | CFSTKVPKRCVNTQPGFCHLCPKPPRYRGNOQVGVGLEAAKTRQVCEPENCKDKTHNCH | 660 | | |
| QY | 601 | CFSTKVPKRCVNTQPGFCHLCPKPPRYRGNOQVGVGLEAAKTRQVCEPENCKDKTHSCH | 660 | | |
| DB | 601 | CFSTKVPKRCVNTQPGFCHLCPKPPRYRGNOQVGVGLEAAKTRQVCEPENCKDKTHSCH | 660 | | |

| | | | | | |
|------------|---|--|------|----------|--|
| QY | 661 | KHAEIYLGHFSDPMYKBCQCTGYAGDGLICGSDSLDGRPNLNLVCATNATYHCICDNC | 720 | | |
| DB | 661 | KHAEIYLGHFSDPMYKBCQCTGYAGDGLICGSDSLDGRPNLNLVCATNATYHCICDNC | 720 | | |
| QY | 721 | PHLPNSGQEDFDKDGIGDADDDDDNDGVDYDEKNCQLLFPNQADYDKDEVDGDCNCP | 780 | | |
| DB | 721 | PKLPNSGQEDFDKDGIGDADDDDDNDGVDYDEKNCQLLFPNQADYDKDEVDGDCNCP | 780 | | |
| QY | 781 | YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQDRTDGDGVDGDCNCP | 840 | | |
| DB | 781 | YVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPYVYNTDQDRTDGDGVDGDCNCP | 840 | | |
| QY | 841 | VHNPQDQTDVNDLVGQDCNNEDIDDDGHQNNODNCPYISNANOADHRDQGGDADPDD | 900 | | |
| DB | 841 | MHNPQDQTDVNDLVGQDCNNEDIDDDGHQNNODNCPYISNANOADHRDQGGDADPDD | 900 | | |
| QY | 901 | DNDGVDDRDNCRLVFNPOEDLDGGRGDIKDDDFDNDNIPDIDDDVCPENNAITDFR | 960 | | |
| DB | 901 | DNDGVDDRDNCRLVFNPOEDLDGGRGDIKDDDFDNDNIPDIDDDVCPENNAITDFR | 960 | | |
| QY | 961 | NFQMVPLDPKGTQIDPNVIRHQKELVQTANSDFGIAVGFDFGSGVDFSGTFYVNTDR | 1020 | | |
| DB | 961 | NFQMVPLDPKGTQIDPNVIRHQKELVQTANSDFGIAVGFDFGSGVDFSGTFYVNTDR | 1020 | | |
| QY | 1021 | DDYAGVFGYQSSSRFYVVMKQVTTQTYWEDQTPRAYGYSGLVKVNNSTTGTGEHLN | 1080 | | |
| DB | 1021 | DDYAGVFGYQSSSRFYVVMKQVTTQTYWEDQTPRAYGYSGLVKVNNSTTGTGEHLN | 1080 | | |
| QY | 1081 | ALWHTGNTQCVRTLWHDPRNIGWKDYATRWHLTHRPKTYIRLVHHEGKQVMAADSGPI | 1140 | | |
| DB | 1081 | ALWHTGNTQCVRTLWHDPRNIGWKDYATRWHLTHRPKTYIRLVHHEGKQVMAADSGPI | 1140 | | |
| QY | 1141 | YDQTYAGGRGLGFLVFSQEMVYFSDLYKECRD | 1171 | | |
| DB | 1141 | YDQTYAGGRGLGFLVFSQEMVYFSDLYKECRD | 1171 | | |
| RESULT 4 | | | | | |
| TSP2_CHICK | | | | | |
| ID | TSP2_CHICK | STANDARD; | PRT; | 1178 AA. | |
| AC | P35440; | | | | |
| DT | 01-JUN-1994 (Rel. 29, Created) | | | | |
| DT | 01-JUN-1994 (Rel. 29, Last sequence update) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | |
| DE | Thrombospondin 2 precursor. | | | | |
| GN | THBS2 OR TSP2. | | | | |
| OS | Gallus gallus (Chicken). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; | | | | |
| OC | Gallus. | | | | |
| OX | NCBI_TaxID=9031; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=91217026; PubMed=2022631; | | | | |
| RA | Lawler J., Duquette M., Ferro P.; | | | | |
| RT | "Cloning and sequencing of chicken thrombospondin.;" | | | | |
| RL | J. Biol. Chem. 266:8039-8043(1991). | | | | |
| CC | -/- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN. | | | | |
| CC | -/- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED. | | | | |
| CC | -/- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY. | | | | |
| CC | -/- SIMILARITY: CONTAINS 1 VWFC DOMAIN. | | | | |
| CC | -/- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS. | | | | |
| CC | -/- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS. | | | | |
| CC | -/- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial | | | | |

Oy 1140 IYDTYAGGRLGLFVFSQEWYFSDLKYECD 1171
Db 1146 IYDTTFAAGRLGLFVFSQEWYFSDLKYECD 1177

RESULT 5
TSPL_HUMAN STANDARD; PRT; 1170 AA.
AC P07996;
DT 01-AUG-1988 (Rel. 08, Last Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THS1 OR TSPI OR TSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Endothelial cells;
RX MEDLINE=87057617; PubMed=2430973;
RA Lawler J., Hynes R.O.;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologies with several different
RT proteins.";
RT J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
RA Baumgaertel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region.";
RT J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RA Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins.";
RT Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RA Laherty C.D., Gierman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription.";
RT J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RT Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CARBOHYDRATE-LINKAGE SITES W-385; S-394; W-438; W-441; T-450; W-498
AND T-507.
RC TISSUE-platelet;
RX PubMed=11067851;
RA Hofsteenge J., Huwiler K.G., Macek B., Hess D., Lawler J.,
RA Mosher D.F., Peter-Katalinic J.;
RT "C-mannosylation and O-fucosylation of the thrombospondin type 1
RT module.";
RT J. Biol. Chem. 276:6485-6498(2001).
CC -/- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and

cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
laminin, type V collagen and integrins alpha-V/beta-1, alpha-
V/beta-3 and alpha-IIB/beta-3.
-/- SUBUNIT: Homotrimer; disulfide-linked.
-/- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
-/- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
-/- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
-/- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
-/- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M25631; AAA36741.1; -
CC EMBL; X04665; CAA28370.1; -
CC EMBL; X14787; CAA32889.1; -
CC EMBL; J04835; AAA61178.1; -
CC EMBL; M99425; AAB59366.1; -
CC PIR; A05172; A05172;
CC PIR; A25812; A25812;
CC PIR; A26155; A26155;
CC PIR; A30140; A30140;
CC PIR; A34274; A34274;
CC HSSP; P35555; IEMN.
CC GlycoSuiteDB; P07996; -
CC MM; 188060; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSPI.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; VWFC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF02412; tsp_3; 8.
CC Pfam; PF00093; vwc; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00209; TSP1; 3.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00214; VWC; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS50092; TSPI; 3.
CC PROSITE; PS01208; VWFC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT SITE 928 928
FT DISULFID 270 270
THROMBOSPONDIN 1.
HEPARIN-BINDING (POTENTIAL).
VWFC.
TSP TYPE-1 1.
TSP TYPE-1 2.
TSP TYPE-1 3.
EGF-LIKE 1.
EGF-LIKE 2.
EGF-LIKE 3.
TSP TYPE-3 1.
TSP TYPE-3 2.
TSP TYPE-3 3.
TSP TYPE-3 4.
TSP TYPE-3 5.
TSP TYPE-3 6.
TSP TYPE-3 7.
C-TERMINAL.
CELL ATTACHMENT SITE (POTENTIAL).
INTERCHAIN (PROBABLE).

```
FT DISULFID 274 INTERCHAIN (PROBABLE).
FT DISULFID 551 BY SIMILARITY.
FT DISULFID 572 BY SIMILARITY.
FT DISULFID 575 BY SIMILARITY.
FT DISULFID 592 BY SIMILARITY.
FT DISULFID 599 BY SIMILARITY.
FT DISULFID 620 BY SIMILARITY.
FT DISULFID 650 BY SIMILARITY.
FT DISULFID 657 BY SIMILARITY.
FT DISULFID 678 BY SIMILARITY.
FT CARBOHYD 248 N-LINKED (GLCNAC. .).
FT CARBOHYD 360 C-LINKED (GLCNAC. .). (POTENTIAL).
FT CARBOHYD 385 C-LINKED (MAN).
FT CARBOHYD 394 O-LINKED (FUC. .).
FT CARBOHYD 438 C-LINKED (MAN).
FT CARBOHYD 441 C-LINKED (MAN).
FT CARBOHYD 450 C-LINKED (FUC. .).
FT CARBOHYD 498 C-LINKED (MAN).
FT CARBOHYD 507 C-LINKED (FUC. .).
FT CARBOHYD 708 N-LINKED (GLCNAC. .). (POTENTIAL).
FT CARBOHYD 1067 N-LINKED (GLCNAC. .). (POTENTIAL).
FT CONFLICT 84 T -> A (IN REF. 2, 3 AND 4).
FT CONFLICT 523 T -> A (IN REF. 2).
SQ SEQUENCE 1170 AA; 129412 MW; 69B3E5AE3A395E CRC64;
```

Query Match 64.0%; Score 4228; DB 1; Length 1170;
Best Local Similarity 61.6%; Pred. No. 5.2e-257; Indels 14; Gaps 6;
Matches 725; Conservative 170; Mismatches 267;

```
QY 1 MVRL-VLLALWMPSTQAGHDKD-TTFDLFSINIRKTIKAKQFRGPDGVPAYRFV 58
DB 3 LAGLVFLMHVCGTNRIPESGDSNVDFELTGAARKSGRRLVKGPDSPAPFIE 62
QY 59 RFDYIPPVNADLSKTKIMRQEGFLTAQLKQDKSGRTLLALGPGLSQRQFIVSN 118
DB 63 DANLIPVPDDKFDQLVDVARTKEGFLLLASLQMKRTKRTLLALERKDHSGQVFSVN 122
QY 119 GPADTLDTLWIDTRHVSLEQVLADSQKNVTVOVAGETVSLHVGCDLISFALDEP 178
DB 123 GKAGTLDLSLTVOGKHVSVSEALLATGOWKSITLFEQEDRAQLYIDCEKMEALDVP 182
QY 179 F---YEHLQAEKSMYVAKGSAHSHFRGLLQNVHLVFNSEVIEDILSKKCGQGGAEIN 235
DB 183 IQSVFTRLASIALRIAKGVND-NFQGVQLQNVRFVFTPTBILRNKGCSSSTSVLLT 241
QY 236 AISENTELRGLPHVTVTEYVGPSSRRPVCERSCELGNNVQELSGHLVLYNPENLK 295
DB 242 L--DNNVNGVSSPAIRTNVIGHKTKDLQATCGISCDLSWLELRLGRTVITVTLQDSIR 299
QY 296 RVSNQDFWELIGGPKTRNMSACWDGRFAENETWVYDSCCTCTCKKFTKICHIQITC 355
DB 300 KYTEENKELANELRRP-----LCYHNGVOYRNNEWTVDSCTECHQNSVYICKVSC 353
QY 356 PPATCASPFVEGECPCSLSHVDGEGWSPNAEWTCQSVTCGSGTQQRGRSCDVTSMTC 415
DB 354 PIMPENATVPDCECCPRCWPSDSADGWSWSEWTSCSTSCGNGIQQRGRSCDSLNNRC 413
QY 416 LGPSIQTRACLSKCDTRTDGGSWHSWSPWSSCSVTCGVGNITRILCNLSPVPMGKN 475
DB 414 EGSSVQTRTHIQECKRQKQDGWHSWSPWSSCSVTCGVDGVTIRILCNLSPVPMGKN 473
QY 476 CKSGGRTRACQACAPCIDRWNSPWPWSACTVTCAGGIRFRVRCNSPEPYGGKACV 535
DB 474 CEGARETRACKDACPINGHGWPNFWDICSVTCGGVQKRSRLCNLTPTQFGGKDCVG 533
```

```
QY 536 DYQEROMCNKRSQPVDCGLSNPCFPGAQCSSRPPDGSSWQSCPVGFLNGHTCEDLDECA 595
DB 534 DVTENOICNKQDCPIDGCLSNPCFAGKCTSTPDGSKWCKACPPGSGNGIQCTDVECK 593
QY 596 LVPDICFSTSKVRCVNTQGFHCLPCPPRYRGNQPVGVGLEAAKTEKQVCEPENCKDK 655
DB 594 EYVDAFCNHNHRCENTDPGYNCLPCPPRFTGSGQFVGQVEHATANKQVCKRNPCTDG 653
QY 656 THNCHKHAECIYLGHFSDPMYKCECQTGAGDGLICGEDSLDGLWNLVLCATNATYHC 715
DB 654 THDCNKNKACNYLGHYSDPMYRCEKPGYAGNGIICGEDTDLGWPNEHLVVCANATYHC 713
QY 716 IKDNCPLHNSGOEDFDKDGIGDACDDDDNDGVTDEKNCOLLFNPRADYDKDEVGR 775
DB 714 KDCNCPNLSQGEDYDKDGIGDACDDDDNDKIPDRONCPHYNPAQYDYDRDDVGR 773
QY 776 CNDPYYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNDCPPYYVNTDQDRDGDGVGDHC 835
DB 774 CNDPYYNHPDQADTDNNGEGDACAADIDGDLNLERNDCQYYVNYVDORDTMDGVGDQC 833
QY 836 DNCPLVHNPQDQTDVNDLVGDCCDNNEDIDDGCHQNNQDNCPPYISNANOADHRCOGDA 895
DB 834 DNCPLHNPQDQDSDSDRIGDTCDDNNDQIDEDGHQNNLNCPPYVPPNANOADHRCOGDA 893
QY 896 CDPDDNDGVPDDRDNCRLVFNPDQEDLDGDRGDKDDFNDNIPDIDDVCPENNAIS 955
DB 894 CDHDDNDGIPDDKONCRLVFNPDQKSDGDRGDKDDFNDNIPDIDDVCPENNAIS 953
QY 956 EYDFRNFQWVLDPKGTTQIDPNWVIRHOGKELVQTANSDPGIAVCFDEGSDVSGTFY 1015
DB 954 EYDFRNFQWVLDPKGTTQIDPNWVIRHOGKELVQTANSDPGIAVCFDEGSDVSGTFY 1013
QY 1016 VNTDRDDYAGVFGYSSSRFVVMKQVTOYVWEDQPTRAYGSGVSLKVVNSTTGTG 1075
DB 1014 INTERDDYAGVFGYSSSRFVVMKQVTOYVWEDQPTRAYGSGVSLKVVNSTTGTG 1073
QY 1076 EHLRNALWHTGNTPGQVRLWHDPRNIGWKDYATYRHLTHRPKTYIRVLVHEGQVMA 1135
DB 1074 EHLRNALWHTGNTPGQVRLWHDPRNIGWKDYATYRHLTHRPKTYIRVLVHEGQVMA 1133
QY 1136 DSGPIYDQYAGRGLGLFVFSQEMVYFSDLYKCECD 1171
DB 1134 DSGPIYDQYAGRGLGLFVFSQEMVYFSDLYKCECD 1169
```

RESULT 6
TSPL_BOVIN
ID TSPL_BOVIN STANDARD; PRT; 1170 AA.
AC Q28178; Q28179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL OR TSP-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLSTEIN; TISSUE-Tooth;
RX MEDLINE=98173773; PubMed=9507054;
RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
Inoue H.;
RT "cDNA cloning of bovine thrombospondin 1 and its expression in
odontoblasts and preodontoblasts";
RL Biochim. Biophys. Acta 1382:17-22(1998).
RN [2]
RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC TISSUE-Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

| | | | | | |
|-----|--|----|------|--|------|
| DR | InterPro: IPR000561; EGF-like. | QY | 114 | EIVSNGPADTLDLTWIDGTRHWVLEVDGLADSQWKNVTVOVAGETYSLHVGCGLIDSF | 173 |
| DR | InterPro: IPR001881; EGF Ca. | Db | 121 | SLISNGRATLDSLGERKQVSVSEDAVLATGWTNTITLFQEDRAQLYVGCNMENA | 180 |
| DR | InterPro: IPR001791; Laminin G. | QY | 174 | ALDEP-----FYEHQAQSKRMVYAKGARESHRGLLQNVHLVFNSEVEDILSKGCGQ | 229 |
| DR | InterPro: IPR000884; TSPN. | Db | 181 | ELDVPQIKIFTENL-ASTAHLRAKGVKVD-NFQGVQLQNVRFVFTTLEAILRNKGLSM | 238 |
| DR | InterPro: IPR003129; TSPN. | QY | 230 | QCAEI--NAISENTETLRLGPHVTTEYVGPSSRRPEVCERSCEBGLNMVDELGLHYL | 286 |
| DR | InterPro: IPR001007; VWFC. | Db | 239 | TNSVITLONPVNGSS-----PAIRNTYIGHKTKDLQAVCGFSCDDLKFLFAEMKGLRTL | 292 |
| DR | Pfam: PF00008; EGF; 2. | QY | 287 | VNOPSENLRKRVNDNQFLWELIGGPPKTRNMSACQDQGRFFAENETWVYDSTCTCKRF | 346 |
| DR | Pfam: PF02210; TSPN; 1. | Db | 293 | VTTLKDOVTKETEKNELIAIV----TRTPGVCLHNGVLHKNRDEWTVDSCTECTQNS | 347 |
| DR | Pfam: PF00090; tsp_1; 3. | QY | 347 | KTICHOITCPCPAYCASPSFEVEGECPCSLHSDVDEGEGSWAEWTQCSVTCSSGTOQRGR | 406 |
| DR | Pfam: PF02412; tsp_3; 8. | Db | 348 | ATICRVSCPLMPCNTATIPDGECCPCWPSADSDWSPWSDMTWPCSVTCVGHGTOQRGR | 407 |
| DR | Pfam: PF00093; VWC; 1. | QY | 407 | SCDVTSTCLGPSIOTRACLSKCDTRIRQDGGSHWSWSSCSTVTCVGNITRIRLNS | 466 |
| DR | SMART: SM00181; EGF; 2. | Db | 408 | SCDSLNNPCGSSVOTRSCQIQCDKRFQDGGSHWSWSSCSTVTCSSGQITRIRLNS | 467 |
| DR | SMART: SM00209; TSP1; 3. | QY | 467 | PYPQMGKCKGSGRETRACQCAPCIDGRWSPWSPWSACTVTCAGGIRERTVRCNSPEP | 526 |
| DR | SMART: SM00210; TSPN; 1. | Db | 468 | PYPQLNGRQCEGEGRENKPCQKPCINGQWGPWSLWDCTVTCGGGMKRLRCLNPKP | 527 |
| DR | PROSITE: PS00022; EGF_1; FALSE_NEG. | QY | 527 | QYGGKACVGVDERQMKRSCPDVGCCLSNPCFPQAQCSFPDGGSWSCSCPCVFLGNGT | 586 |
| DR | PROSITE: PS01186; EGF_2; 1. | Db | 528 | QYEGKDCIGEPDSDQCNKQDCPIDGCLSNPCFAGVKCTSFIDGSKWCSGCPGYRNGI | 587 |
| DR | PROSITE: PS00092; TSP1; 3. | QY | 587 | HCEDLDECALVPDICFSTSKVPCVNTOPGFHCLPCPRYRGNQPVGVGLEAAKTEKQVC | 646 |
| DR | PROSITE: PS01208; VWFC; 1. | Db | 588 | TKCDIDKEKVPDCACTFLNGVHRCENTEPGYNCLPCPRFTGTQPGKIEBAKANKQVC | 647 |
| KW | Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; | QY | 647 | EPENPCDKTKNCHKHAECIYLGHFSDPMYKCECQTGYAGDGLIGEDSLDGLWPNLNV | 706 |
| FT | EGF-like domain; Signal. | Db | 648 | KPRNFCADGTHDCHKNARCIYLGHSYSDPMFRCERCPGYAGNGIICGEDTLDGWPENLT | 707 |
| FT | SIGNAL 1 22 | QY | 707 | CATNATYHCIKCNPLNSGDEDFKQIGIGACDDDDNDGVDTEKQNCQLLPNPROAD | 766 |
| FT | CHAIN 23 1173 | Db | 708 | CVDNATYHCLKQNCPLNSGQEDYDKDGMGACDKDDNDGILDDRDNCQFYVNPAYD | 767 |
| FT | DOMAIN 23 235 | QY | 767 | YDKDEVDGRCDCNCPVYHNPQJDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRT | 826 |
| FT | DOMAIN 319 376 | Db | 768 | YDRDDVDGRCDCNCPVYHNPQJDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRT | 827 |
| FT | DOMAIN 382 433 | QY | 827 | DGDDGVGDHCDNCPVYHNPQJDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRT | 886 |
| FT | DOMAIN 382 433 | Db | 828 | DKDGVGDHCDNCPVYHNPQJDTDNNGEGDACSVDIDGDDVFNERNDCNCPYVNTDQRT | 887 |
| FT | DOMAIN 438 494 | QY | 887 | HRDGGDGDACDDNDGVPDDRDNCRLVFNPDQEDLDGGRGDKDKDDFDNDNIPDID | 946 |
| FT | DOMAIN 495 546 | Db | 888 | HRDGGDGDACDDNDGVPDDRDNCRLVFNPDQEDLDGGRGDKDKDDFDNDNIPDID | 947 |
| FT | DOMAIN 550 590 | QY | 947 | VCPENNAISETDFRNFQVPLDPKGTQIDPNWVIRHQKELVQVANSQDGLAVGFDEFG | 1006 |
| FT | DOMAIN 591 648 | Db | 948 | VCPENVEISTTDFRNFQVPLDPKGTQIDPNWVIRHQKELVQVANSQDGLAVGFDEFG | 1007 |
| FT | DOMAIN 649 693 | QY | 1007 | SVDFSGTFVYNTDRDDYAGFVFGYQSSSRFYVVMKQVYQTYWEDQPTRAYGSGVSK | 1066 |
| FT | DOMAIN 726 761 | Db | 1008 | AVDFSGTFVYNTDRDDYAGFVFGYQSSSRFYVVMKQVYQTYWEDQPTRAYGSGVSK | 1067 |
| FT | DOMAIN 762 784 | QY | 1067 | VYNTSTGTGEHLRNALWHTGNTPGQVRLTWHDPNRIGKDYATYRHLHTRPKTGIRVL | 1126 |
| FT | DOMAIN 785 820 | Db | 1068 | VYNTSTGTGEHLRNALWHTGNTPGQVRLTWHDPNRIGKDYATYRHLHTRPKTGIRVL | 1127 |
| FT | DOMAIN 821 843 | QY | 1127 | VHEGQVADSGPIYDQTVAGGRLGLFVFSQEMVYFSDLYKECRD | 1171 |
| FT | DOMAIN 844 881 | Db | 1128 | MYEGKRVWADSGPIYDQTVAGGRLGLFVFSQEMVYFSDLYKECRD | 1172 |
| FT | DOMAIN 882 917 | | | | |
| FT | DOMAIN 918 953 | | | | |
| FT | DOMAIN 954 1173 | | | | |
| FT | SITE 929 931 | | | | |
| FT | DISULFID 554 565 | | | | |
| FT | DISULFID 559 575 | | | | |
| FT | DISULFID 578 589 | | | | |
| FT | DISULFID 595 611 | | | | |
| FT | DISULFID 602 620 | | | | |
| FT | DISULFID 623 647 | | | | |
| FT | DISULFID 653 666 | | | | |
| FT | DISULFID 660 679 | | | | |
| FT | DISULFID 681 692 | | | | |
| FT | CARBOHYD 155 155 | | | | |
| FT | CARBOHYD 158 158 | | | | |
| FT | CARBOHYD 250 250 | | | | |
| FT | CARBOHYD 363 363 | | | | |
| FT | CARBOHYD 705 705 | | | | |
| FT | CARBOHYD 711 711 | | | | |
| FT | CARBOHYD 1070 1070 | | | | |
| SEQ | SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64; | | | | |

Query Match 63.4%; Score 4189; DB 1; Length 1173;
 Best Local Similarity 61.3%; Pred. No. 1.4e-254; Indels 34; Gaps 9;
 Matches 726; Conservative 161; Mismatches 264;
 QY 5 LVLLALWVWPSQAGHQ-----DKDTDFLSISINRKT-----IGAKQFRGPDGPV 53
 Db 4 IFLLLVMPQT---HQAAESGNDNSVDFLPELTGYNRKAGSRKPGQLHLVKGDPDSSP 60
 QY 54 AYRFVRFYIPPNVADLSKTYKIMQKGEFFLTAQLKQDGKSRGFTLLALEGFGLSQRF 113
 Db 61 AVRIEDADLIPLPEDKFQDLDLAIADRGFILLATLROAKSRGALLSVKRDGGGHIF 120

RESULT 9
COMP_HUMAN

AC P49747; Q16388; STANDARD; PRT; 757 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=95229140; PubMed=7713493;
RA Newton G., Weremowicz S., Morton C.C., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Lawler J.;
RT "Characterization of human and mouse cartilage oligomeric matrix
RT protein.";
RL Genomics 24:435-439(1994).
RN [2]
RP VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.
RX MEDLINE=95400301; PubMed=7670471;
RA Hecht J.T., Nelson L.D., Crowder E., Wang Y., Elder F.F.B.,
RA Harrison W.R., Francomano C.A., Prange C.K., Lennon G.G., Deere M.,
RA Lawler J.;
RT "Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)
RT cause pseudoachondroplasia.";
RL Nat. Genet. 10:325-329(1995).
RN [3]
RP VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.
RX MEDLINE=95400302; PubMed=7670472;
RA Briggs M.D., Hoffman S.M.G., King L.M., Olsen A.S., Mohrenweiser H.,
RA Leroy J.G., Rimo D.L., Lachman R.S., Gaines E.S.,
RA Cekleniak J.A., Knowlton R.G., Cohn D.H.;
RT "Pseudoachondroplasia and multiple epiphyseal dysplasia due to
RT mutations in the cartilage oligomeric matrix protein gene.";
RL Nat. Genet. 10:330-336(1995).
RN [4]
RP VARIANT MED LYS-523.
RX MEDLINE=97173141; PubMed=9021009;
RA Ballo R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,
RA Ramesar R.S.;
RT "Multiple epiphyseal dysplasia, ribbing type: a novel point mutation
RT in the COMP gene in a South African family.";
RL Am. J. Med. Genet. 68:396-400(1997).
RN [5]
RP VARIANT MED SER-371, AND VARIANT PSACH 513-VAL--LYS-516 DEL.
RX MEDLINE=97327574; PubMed=9184241;
RA Susic S., McGroarty J., Ahier J., Cole W.G.;
RT "Multiple epiphyseal dysplasia and pseudoachondroplasia due to novel
RT mutations in the calmodulin-like repeats of cartilage oligomeric
RT matrix protein.";
RL Clin. Genet. 51:219-224(1997).
RN [6]
RP VARIANTS PSACH AND MED.
RX MEDLINE=98130533; PubMed=9463320;
RA Briggs M.D., Mortier G.R., Cole W.G., Golik S.S.,
RA Bonaventura J., Nuytink L., de Paape A., Leroy J.G., Biesecker L.,
RA Lipson M., Wilcox W.R., Lachman R.S., Rimo D.L., Knowlton R.G.,
RA Cohn D.H.;
RT "Diverse mutations in the gene for cartilage oligomeric matrix protein
RT in the pseudoachondroplasia-multiple epiphyseal dysplasia disease
RT spectrum.";
RL Am. J. Hum. Genet. 62:311-319(1998).
RN [7]
RP VARIANTS PSACH AND MED.
RX MEDLINE=99118868; PubMed=9921895;
RA Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Sannohe A.,
RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.;
RT "Novel and recurrent COMP (cartilage oligomeric matrix protein)

RT mutations in pseudoachondroplasia and multiple epiphyseal dysplasia.";
Hum. Genet. 103:633-638(1998).
RN [8]
RP VARIANTS PSACH AND MED.
RX MEDLINE=98112405; PubMed=9452026;
RA Loughlin J., Irven C., Mustafa Z., Briggs M.D., Carr A., Lynch S.A.,
RA Knowlton R.G., Cohn D.H., Sykes B.;
RT "Identification of five novel mutations in cartilage oligomeric
RT matrix protein gene in pseudoachondroplasia and multiple epiphyseal
RT dysplasia.";
RL Hum. Mutat. Suppl. 1:S10-S17(1998).
RN [9]
RP VARIANT PSACH GLY-482.
RX MEDLINE=98112442; PubMed=9452063;
RA Susic S., Ahier J., Cole W.G.;
RT "Pseudoachondroplasia due to the substitution of the highly conserved
RT Asp482 by Gly in the seventh calmodulin-like repeat of cartilage
RT oligomeric matrix protein.";
RL Hum. Mutat. Suppl. 1:S125-S127(1998).
CC -|- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
CC -|- DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDOACHONDROPLASIA
CC (PSACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDM), WHICH ARE
CC DOMINANTLY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT
CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED
CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH
CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.
CC -|- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -|- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L32137; AAA57253.1; -;
CC EMBL; S79499; AAB35289.1; -;
CC EMBL; S79500; AAB35270.1; -;
CC HSSP; P02468; 1KLO.
CC MIM; 600310; -;
CC MIM; 132400; -;
CC MIM; 177170; -;
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF02412; tsp_3; 11.
CC SMART; SM00179; EGF_CA; 2.
CC SMART; SM00001; EGF_like; 2.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01187; EGF_CA; 2.
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
KW Signal; Disease mutation.
FT SIGNAL 1 20
FT CHAIN 21 757
FT DOMAIN 22 86
FT DOMAIN 87 126
FT DOMAIN 127 179
FT DOMAIN 180 222
FT DOMAIN 225 267
FT DOMAIN 297 332
FT DOMAIN 333 355
FT DOMAIN 356 391
FT DOMAIN 392 414
FT DOMAIN 415 452
FT DOMAIN 453 488
FT DOMAIN 489 524
FT DOMAIN 525 757
FT C-TERMINAL 69 69
FT INTERCHAIN (PROBABLE).
FT DISULFID 69 69
FT DISULFID 72 72
FT INTERCHAIN (PROBABLE).
FT DISULFID 72 72

FT DISULFID 91 102 BY SIMILARITY.
FT DISULFID 96 111 BY SIMILARITY.
FT DISULFID 114 125 BY SIMILARITY.
FT DISULFID 131 142 BY SIMILARITY.
FT DISULFID 136 151 BY SIMILARITY.
FT DISULFID 154 178 BY SIMILARITY.
FT DISULFID 184 197 BY SIMILARITY.
FT DISULFID 191 206 BY SIMILARITY.
FT DISULFID 209 221 BY SIMILARITY.
FT DISULFID 229 243 BY SIMILARITY.
FT DISULFID 237 253 BY SIMILARITY.
FT CARBOHYD 255 266 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 367 369 CELL ATTACHMENT SITE (POTENTIAL).
FT VARIANT 290 290 D -> N (IN PSACH; MILD FORM).
FT VARIANT 299 299 /FTID-VAR_007614.
FT VARIANT 328 328 G -> R (IN PSACH; MILD FORM).
FT VARIANT 342 342 /FTID-VAR_007615.
FT VARIANT 349 349 C -> R (IN PSACH; MILD FORM).
FT VARIANT 361 361 D -> Y (IN MED; FAIRBANK TYPE).
FT VARIANT 361 361 D -> V (IN PSACH; MILD FORM).
FT VARIANT 367 368 /FTID-VAR_007617.
FT VARIANT 371 371 D -> V (IN MED; FAIRBANK TYPE).
FT VARIANT 372 372 /FTID-VAR_007620.
FT VARIANT 374 374 MISSING (IN MED).
FT VARIANT 387 387 /FTID-VAR_007621.
FT VARIANT 391 394 C -> S (IN MED; FAIRBANK TYPE).
FT VARIANT 408 408 /FTID-VAR_007622.
FT VARIANT 440 440 /FTID-VAR_007623.
FT VARIANT 440 440 MISSING (IN PSACH).
FT VARIANT 453 453 /FTID-VAR_007624.
FT VARIANT 459 459 G -> E (IN PSACH; MILD FORM).
FT VARIANT 468 468 /FTID-VAR_007625.
FT VARIANT 469 469 G -> R (IN PSACH; MILD FORM).
FT VARIANT 472 472 /FTID-VAR_007626.
FT VARIANT 472 472 D -> Y (IN MED; FAIRBANK TYPE).
FT VARIANT 472 472 /FTID-VAR_007627.
FT VARIANT 472 472 G -> E (IN PSACH; MILD FORM).
FT VARIANT 472 472 /FTID-VAR_007628.
FT VARIANT 472 472 G -> R (IN PSACH; MILD FORM).
FT VARIANT 472 472 /FTID-VAR_007629.
FT VARIANT 472 472 N -> S (IN MED; FAIRBANK TYPE).
FT VARIANT 472 472 /FTID-VAR_007630.
FT VARIANT 472 472 MISSING (IN PSACH; SEVERE FORM).
FT VARIANT 472 472 C -> Y (IN PSACH; SEVERE FORM).
FT VARIANT 472 472 /FTID-VAR_007631.
FT VARIANT 472 472 MISSING (IN PSACH).
FT VARIANT 472 472 /FTID-VAR_007632.
FT VARIANT 472 472 /FTID-VAR_007633.
FT VARIANT 472 472 D -> Y (IN PSACH; SEVERE FORM).
Query Match 30.2% Score 1993; DB 1; Length 757;
Best Local Similarity 54.1% Pred. No. 2.1e-117;
Matches 362; Conservative 73; Mismatches 176; Indels 58; Gaps 6;
QY 549 PVDGCLSNPCPGACQSFPGDGSNCGSPVGLNGTHCEDLDRCALVPDICTFSTSKVP 608
Db 87 PLLHCAPGCFPGVACIQTESGG-RCGCPGAGFTGSHCTDVNRCNAHP--CEPR---V 140
QY 609 RCVNTQPGFHCPLCPPRYRGNOPVGVGLEAAKTEKQVCEPENPCPKDKTHN----- 658
Db 141 RCINTSPGRCACPGYSGTHQGVGLAFKANKQVCTDINECTGHOHCVPSVCLNT 200
QY 659 -----CHKHACVILYGHFSPMYKCEQT 682
Db 201 RGSFOGCPQPGVGDQAGCGRGAORFCPDGSPSECHHADCVL---ERDSRSCVCRV 257

QY 683 GYAGDGLIGEDSDLDGWPNLNLVLCATNATYHCIKDKNCPLHNSGOEDFEDKDGIGDADD 742
Db 258 GWAGNGILCGRDYDLGFFDEKLRCEP---QCRKNCVTPVNSGQEDVDROGIGDADCP 314
QY 743 DDNDGVTDEKNCQLLEFNPRQADYDKDEVDGDCDCNCPVYHNPAQIDTDNNGEGDACSVD 802
Db 315 DADGCGVPNEKNCPLVRNPQQRNTDEKWDGACDNCRSQKNDQDKDTQDQDGRGDACDD 374
QY 803 IDGDDVFNERNDCPVYVNTDQRTDGDGVDGDCDCNCPVHNPDQTDVDNDLVGDQDNE 862
Db 375 IDGDRIRNADNCPRVPSDQSDGIGDADNCPCQKSNPDQADVDHDFVGDACDSQ 434
QY 863 DIDDGHHQNNQNCQPYISNAQADHDDGOGDADCDPDDNDGVPDDRDNCRLVFNPDQED 922
Db 435 DQDGDGHQSDRNCPCVPSAQSDESDHDDGOGDADCDDDNDGVPDSRDNCRLVFNPDQED 494
QY 923 LQDGRGDIKDDDFDNDNIPDIDVPCPENNAISETDFRNFQMVPLDPKTKTQIDPNWVIR 982
Db 495 ADRGSGVDVQDDFADKVDKIDVCPENAEVTLTDFRAFTVVLDPEDGAQIDPNWVIL 554
QY 983 HOGKELVOTANSDDPGTAVGDFDEGSVDFTGVVNTDRDDYAGVFGYQSSRFVVMW 1042
Db 555 NQGREIVQTMNSDPGLAVGITAENGVDFTGTHVNTVDDYAGFIFGYQDSSSFVVMW 614
QY 1043 KQVQTYWEDQPTRAYGYSGLKVVNSTTGTGEHLRNALWHTGNTPPGQVRLWHPDRI 1102
Db 615 KQMEQTYWQANPFRVAEPGIQKAVKSTGEGEQLRNALWHTGNTESQVRLWHPDRI 674
QY 1103 GWKDTATYRWHLTHRPKTYIRVLVHVGQVMDSGPIYDQYAGRLGLFVFSQBMVYF 1162
Db 675 GWKDKSYRWFLQHRPQVGYIRVREYFEGPELVADSNVLDLTTMRGRLGVFCFSQBNIIW 734
QY 1163 SLDKVECD 1171
Db 735 ANLRYCND 743
RESULT 10
ID COMP RAT STANDARD; PRT; 755 AA.
AC P35444;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93054522; PubMed=1429587;
RA Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;
RT "COMP (cartilage oligomeric matrix protein) is structurally related
to the thrombospondins.";
RL J. Biol. Chem. 267:22346-22350(1992).
RN (2)
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.
RX MEDLINE=97020114; PubMed=8864111;
RA Malashkevich V.N., Kammerer R.A., Efimov V.P., Schulthess T.,
RT Engel J.;
RT "The crystal structure of a five-stranded coiled coil in COMP: a
prototype ion channel?";
RL Science 274:761-765(1996).
CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

RT "Thrombospondin 3 (Thbs3); a new member of the thrombospondin gene
 RL family.", Chem. 267:12192-12196(1992).
 RN [4]
 RP SEQUENCE OF 834-956 FROM N.A.
 RC STRAIN-BALB/C;
 RA Vos H.L., Mockingsturm-wilson M., Rood P.M.L., Maas A.C.E.,
 RA Duhig T., Gendler S.J., Bornstein P.;
 RA Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -!- SUBUNIT: OLIGOMER; DISULFIDE-LINKED.
 CC -!- TISSUE SPECIFICITY: BRAIN, LUNGS AND CARTILAGE.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L04302; AAA040497.1; -
 DR EMBL; L24434; AAA040433.1; -
 DR EMBL; M86620; AAA040430.1; -
 DR EMBL; M86611; AAA040430.1; JOINED.
 DR EMBL; M86612; AAA040430.1; JOINED.
 DR EMBL; M86613; AAA040430.1; JOINED.
 DR EMBL; M86614; AAA040430.1; JOINED.
 DR EMBL; M86615; AAA040430.1; JOINED.
 DR EMBL; M86616; AAA040430.1; JOINED.
 DR EMBL; M86617; AAA040430.1; JOINED.
 DR EMBL; M86618; AAA040430.1; JOINED.
 DR EMBL; M86619; AAA040430.1; JOINED.
 DR EMBL; U16175; AAA98537.1; -
 DR PIR; A46016; A46016.
 DR PIR; A44124; A44124.
 DR HSSP; P35555; 1EMN.
 DR MGD; MGI:98739; Thbs3.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR003367; tsp.3.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF02412; tsp.3; 8.
 DR SMART; SM00179; EGF_Ca; 2.
 DR SMART; SM00001; EGF_like; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_Ca; 2.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 956
 FT DOMAIN 22 272
 FT DOMAIN 272 315
 FT DOMAIN 316 369
 FT DOMAIN 370 413
 FT DOMAIN 414 456
 FT DOMAIN 488 523
 FT DOMAIN 524 546
 FT DOMAIN 547 582
 FT DOMAIN 583 605
 FT DOMAIN 606 643
 FT DOMAIN 644 683
 FT DOMAIN 684 719
 FT DOMAIN 720 956
 FT DISULFID 266 266
 INTERCHAIN (PROBABLE).

FT DISULFID 269 269 INTERCHAIN (PROBABLE).
 FT DISULFID 278 289 BY SIMILARITY.
 FT DISULFID 283 300 BY SIMILARITY.
 FT DISULFID 303 314 BY SIMILARITY.
 FT DISULFID 320 332 BY SIMILARITY.
 FT DISULFID 326 341 BY SIMILARITY.
 FT DISULFID 344 368 BY SIMILARITY.
 FT DISULFID 374 388 BY SIMILARITY.
 FT DISULFID 382 397 BY SIMILARITY.
 FT DISULFID 400 412 BY SIMILARITY.
 FT DISULFID 418 432 BY SIMILARITY.
 FT DISULFID 426 442 BY SIMILARITY.
 FT DISULFID 444 455 BY SIMILARITY.
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 937 937 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 433 433 L -> F (IN REF. 2).
 FT CONFLICT 563 563 N -> D (IN REF. 2).
 FT CONFLICT 720 720 G -> E (IN REF. 2).
 FT CONFLICT 871 871 L -> W (IN REF. 2 AND 4).
 SQ SEQUENCE 956 AA; 103972 MW; BFFA666488C069A2 CRC64;
 Query Match 29.3%; Score 1936.5; DB 1; Length 956;
 Best Local Similarity 52.0%; Pred. No. 9.4e-114;
 Matches 359; Conservative 81; Mismatches 181; Indels 69; Gaps 11;
 QY 537 VDERQMC---NKRSCPVDGCLSNPCFPFQAQCS---SFPDGSWSGSCPVPGLNGTHCED 590
 DB 263 IMECQVCGHEQRS-----HCSPPSPCERGVDCMEVIEYP--GYRCGCPGLQNGTHCDD 316
 QY 591 LDECALVPDICTFSTKVPKVCNTQPGFHLCPKPPRYGNGQVPGVLEAAKTEKQVCE--- 647
 DB 317 INECAHA-DPCFPSS---CINTMPGFHCEACPPYKGTGTVSGVGDYARASQVNCID 372
 QY 648 -----PENCKDKTHN-CHKHAE 665
 DB 373 ECNDGNGGCDPNISICTNTVSGFKGCPGLGNGSQGCVPARTCHSPAHSCHHAHC 432
 QY 666 IYLGHSFDPWKCECOTGYAGDLIGEDSDLDGWNPLNLVCATNATYHCKNCPHLPN 725
 DB 433 LF---ERNGAVSCOCNVHAGNVCVGPDTIDGYDQALPCMDN-NKHCQDNCLLTPN 488
 QY 726 SGOEDFDKDGIGDADDDDDNDGVTDEKQNCQLLFPNRPADYDKDEVGRDCNCPVHNP 785
 DB 489 SGOEDADNDGVGQDQDDADGDIKNVEDNCRFLPNKQDQNSDTSFGDACDCNCPVNN 548
 QY 786 AQIDTDNNGEGDACSVDIDGDDVFNERNDCPVYNTDQRTDGDGVDGHCNCPVHNP 845
 DB 549 DQKDTDNGEGDACCNDNDVDGDIPLNGLDNCPKVPNPLQTDREDDGVDGACDCSPMSNPT 608
 QY 846 QTDVDNDLVGDQDNNEDIDDDGHQNNQNCVPYISNANQADHRDQGGDADCDPDDNDGV 905
 DB 609 QTDADSDLVGDCVTNEDSDSGDHQDTKNCQPLPNSQLSDNDGLGDECDGDDNDGV 668
 QY 906 PD-----DRNCRILVFPNDQEDLDGDRGDIKDDDFNDNIPDIDDDVCPENNALSETFRN 961
 DB 669 PDYIPGPNCRILVFPNPKDSDGNGVGDVCEDDFNDVAVDPLDYCPESAGVTLTDFRA 728
 QY 962 FQWVPLDPKTTQIDPNWVIRHOGKELVQTSANDPCIAVGDFGSGVDFSGTGYVNTDRD 1021
 DB 729 YQTVILDPGGDAQIDPNWVVLNQGMIEVQTMNSDPLAVGYTAFNGVDFEGTFHVTYTD 788
 QY 1022 DDYAGVFYGYSSSFYVVMKQVQTYTWEDQPTRAYGYSGLSKVYVNSTGTGHLRNA 1081
 DB 789 DDYAGLFYSQDSGREYVVMKQTEQTYWQATPFRAVAPGLKAVTSISGSGHLRNA 848
 QY 1082 LWHGTGNTPGQVRLWHDPRNIGWKDYTAIRWLTHRPKGTGYRLVHESQGWADSGPIY 1141
 DB 849 LWHGTGTPDQVRLWLDPRNVGLRDKTSYRWRLHRRPQVGYIRVKLYEGPQLVADSGVYI 908
 QY 1142 DQTYAGRLGLFVFSQEMVYFSDLYKVECD 1171

Db 929 IWSNLKYRCND 939

RESULT 14

TSP4_HUMAN STANDARD; PRT; 961 AA.

AC P35443;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Thrombospondin 4 precursor.

GN THBS4 OR TSP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

ON NCBI_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=93353522; PubMed=9350346;

RA Lawler J., Duquette M., Uffry L., McHenry K., Smith T.F.;

RT "The evolution of the thrombospondin gene family.";

RL J. Mol. Evol. 36:509-516(1993).

CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.

CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.

CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; Z19585; CAA79635.1; -.

DR PIR; S36069; S36069.

DR HSSP; P02468; 1KLO.

DR TM; 600715; -.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF_Ca.

DR InterPro: IPR003129; TSPN.

DR InterPro: IPR003367; tsp_3.

DR Pfam; PF02210; TSPN; 1.

DR Pfam; PF02412; tsp_3; 9.

DR SMART; SM00179; EGF_CA; 2.

DR SMART; SM00001; EGF-like; 2.

DR SMART; SM00210; TSPN; 1.

DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 2.

KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;

KW Signal.

FT SIGNAL 1 21

FT CHAIN 22 961

FT DOMAIN 22 285

FT DOMAIN 286 325

FT DOMAIN 326 378

FT DOMAIN 379 419

FT DOMAIN 420 462

FT DOMAIN 492 527

FT DOMAIN 528 550

FT DOMAIN 551 586

FT DOMAIN 587 609

FT DOMAIN 610 647

FT DOMAIN 648 687

FT DOMAIN 688 723

FT DOMAIN 724 961

FT SITE 562 564

POTENTIAL.

THROMBOSPONDIN 4.

N-TERMINAL.

EGF-LIKE 1.

EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 4.

TSP TYPE-3 1.

TSP TYPE-3 2.

TSP TYPE-3 3.

TSP TYPE-3 4.

TSP TYPE-3 5.

TSP TYPE-3 6.

TSP TYPE-3 7.

C-TERMINAL.

CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 258 258 INTERCHAIN (PROBABLE).

FT DISULFID 261 261 INTERCHAIN (PROBABLE).

FT DISULFID 290 301 BY SIMILARITY.

FT DISULFID 295 310 BY SIMILARITY.

FT DISULFID 313 324 BY SIMILARITY.

FT DISULFID 330 341 BY SIMILARITY.

FT DISULFID 335 350 BY SIMILARITY.

FT DISULFID 353 377 BY SIMILARITY.

FT DISULFID 383 394 BY SIMILARITY.

FT DISULFID 388 403 BY SIMILARITY.

FT DISULFID 406 418 BY SIMILARITY.

FT DISULFID 424 438 BY SIMILARITY.

FT DISULFID 432 448 BY SIMILARITY.

FT DISULFID 450 461 BY SIMILARITY.

FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 961 AA; 105801 MW; 55A48AF481AEB9DD CRC64;

Query Match 28.6%; Score 1890; DB 1; Length 961;

Best Local Similarity 53.1%; Pred. No. 7.7e-111;

Matches 355; Conservative 71; Mismatches 185; Indels 58; Gaps 9;

QY 549 PVDGCLSNPCFPGAQCSFPDGSWSCSCPVGFLNGTHCEDLDECALVPDICESTSKVP 608

Db 286 PPRCDSNPFCRGVQCSTDSRDG-FQCGPCEGYTGNGTICIDVDECKYHP--CYGP--V 339

QY 609 RCVNTOPGFHCLPCPPRYRGNQPVGLEAAKTEKQVC-----EPEN----- 650

Db 340 HCINLSFGFRCDACPVFTGPMVQGVGISFAKSNKVQCTIDICRNGACVPNSICVNLG 399

QY 651 -----PCK-----DKTHNCHKHAEIYLGHFSDPMYKCEQCTGYAG 686

Db 400 SYRCGPKCPGYTGQIRGCKVERNCRNPELNPQSVNAQCI---EERQGDVTCVCGVGWAG 456

QY 687 DGLICEDSLDGLWPNLNLVCATNATYHCKDKNCPLHPSNGSEDFDKDGLGACDDDDN 746

Db 457 DGTICGKVDIDISYDEELPCFSAR---NCKKDKCYVPSGQEDADRDLGIGDACEADG 513

QY 747 DGYTDEKDNQQLFNPRQADYDDEVGDRDCNCPYVHNPAQIDTDNNGEGDACSVDIDG 806

Db 514 DGIILNEQDNCVLIHNVDQRNSDKDIFGDACDNCLSVLNNDQKDTDGDGRGDACDDMDG 573

QY 807 DVNERDNCPIVYNTDQRTDGDGVGDHDCNCPVLHNPQDQTDVNDLVGDCDNNEDID 866

Db 574 GIKNILDNCPKFPNRQDKDGDGVGDACDSCPVSNPQSDVDNDLVGDCDNTQDSG 633

QY 867 DGHONNODNCPYISNANOADHRDGGDGDACDPPDDNDGVPD---DRDNCRLVFNPDQED 922

Db 634 DGHQDSTDNCPVINSQAQLDTRDKDGDGDECDDDDDNDGIFDLVPPGPDNCRVLPNPAQED 693

QY 923 LDGDRGDCICKDDFDNDNIPDIDDVCPENNAISETDFRNFQMVLPDPKGTQIDPNVIR 982

Db 694 SNSDGVGDICESDFDQDQVIDRIDVCPENAEVLTDFRAYQTVGLDPEGDAQIDPNVVL 753

QY 983 HQGKELVQTANSPDGTAVGDFEGSVDFTFYVNTDRDDYAGVFGYSSSRFYVMW 1042

Db 754 NQGMIEVQTMNSDPLGAVGYTAFFNGVDFEFTFVNTQDDYAGVFGYSSSFYVMW 813

QY 1043 KOVTVYWEQDPTFRAYGVSILKVVNSTGTGTEHLRNALWHTGTPGVRTLWHDPRNI 1102

Db 814 KQEQYVQATPPRAVAPGQIKAVKSKTGPGEHLRNSLWHTGTDSDQVRLWKSQNV 873

QY 1103 GWKDYATYRWHLTHRPKTYGIRVLVHEGKQVMSADSPYDQTYAGRGLGLFVFSQEMVYF 1162

Db 874 GWKDKYSYRWFLQHRPQGVYIRVRYEGSELVADSGVTIDTTRGGRGLGVCFSQENIIV 933

QY 1163 SLDKYECD 1171

Db 934 SNLYRCND 942

RESULT 15

TSP4_RAT
ID TSP4_RAT STANDARD; PRT; 980 AA.
AC P49744;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 4 precursor.
GN THBS4 OR TSP4 OR TSP-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=LEWIS; TISSUE=Skeletal muscle;
RX MEDLINE=96074771; PubMed=7490284;
RA Arber S.; Caroni P.;
RT "Thrombospondin-4, an extracellular matrix protein expressed in the
developing and adult nervous system promotes neurite outgrowth.";
RL J. Cell Biol. 131:1083-1094(1995).
CC - FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC - SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS (BY
SIMILARITY).
CC - SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC - SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X89963; CAA62002.1;
DR HSP: P35444; LVDF.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR003367; tsp_3.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF02412; tsp_3; 9.
DR SMART: SM00179; EGF_CA; 2.
DR SMART: SM00001; EGF_Like; 2.
DR SMART: SM00210; TSPN; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 2.
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
Signal.
FT SIGNAL 1 39
FT CHAIN 40 980
FT DOMAIN 40 303
FT DOMAIN 304 343
FT DOMAIN 344 396
FT DOMAIN 397 437
FT DOMAIN 438 481
FT DOMAIN 511 546
FT DOMAIN 547 569
FT DOMAIN 570 605
FT DOMAIN 606 628
FT DOMAIN 629 666
FT DOMAIN 667 706
FT DOMAIN 707 742
FT DOMAIN 743 980
FT DISULFID 276 276
FT DISULFID 279 279
FT DISULFID 308 319
FT DISULFID 313 328
FT DISULFID 331 342

FT DISULFID 348 359 BY SIMILARITY.
FT DISULFID 363 368 BY SIMILARITY.
FT DISULFID 371 395 BY SIMILARITY.
FT DISULFID 401 412 BY SIMILARITY.
FT DISULFID 406 421 BY SIMILARITY.
FT DISULFID 424 436 BY SIMILARITY.
FT DISULFID 442 456 BY SIMILARITY.
FT DISULFID 450 466 BY SIMILARITY.
FT DISULFID 468 480 BY SIMILARITY.
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 980 AA; 108213 MW; 056D41EB6E206FCF CRC64;

Query Match 28.5%; Score 1881; DB 1; Length 980;
Best Local Similarity 36.8%; Pred. No. 2.9e-110; Indels 290; Gaps 32;
Matches 440; Conservative 124; Mismatches 341;

Qy 6 VLLALWVWSPSTQAGHQDKDTTFDLFSININRKTIGAKQFRGDPFGVPAYREVFDYIP 65
Db LLLHLVLPQWQAGAAATPQVFDLLP-SSQRLNPAALQ-----P 66

Qy 66 VNADLSKITIMRQKEGFFLTAQLKQDGKSRGTLLEGLPGLSORQFIVSNGPADTLD 125
Db VLTDP-----TLHELVIISTFKLSKSSATIFGLYSSSDNSKYFEFTVMGRINKAI 117

Qy 126 LTYW-IDGTRVHVSLEDVGLADSOAKNV-----TVQVAGETYSLVHVGCDLIDSF-ALDEP 178
Db LRYLKKDDGKIHLVFNNNQLADGRHRLRLSLNSLRQAGSVELYLDQVQVSVNNLPRA 177

Qy 179 F---YEHLOAKSRMYVARGSARESHFRGLLQNVHLVFNVSVEDILSKKCGOQGAEN 235
Db FSGLTQNPQAIELRTFFQRP-----QDFLEELKLVVRSLSFQVASLQDCFLQOSEPLA 230

Qy 236 AISENTETLRGLPHVTTEYVGPSSRRPEVCERSEEELGNMVELSGLHLVNLSENK 295
Db 231 A-----TGTGDFNR-----QFLGQMTQ-----LNLLEGEYKDLLR 260

Qy 296 RVSNDNQFLWELGGPKTRNMSACWDGRFAENETWVVDSTCTCTCKKFTTCHQITC 355
Db 261 QOVKTSFL-----RN-----TIAECQAC 279

Qy 356 PPATCASPFEVEGECPCSLHSVDGEGSWPAEWTOCSVTQSGTQQRGSCDVTSTNTC 415
Db 280 GPLSFQSP-----PNTLVPI-----APPAPTRTRRCDSSPCFRGVR----- 319

Qy 416 LGPSIQTRACSLSKCDTRIRQDGGHSHSPSSCSVTGCVGNITRILCNLSPVPMGGKN 475
Db 320 -----TDTR-----DG-----FQCG----- 329

Qy 476 CKSGRETKACQAGPCPIDGRWSPWSPSACTVTTCAGGIRTRVCNSPEPOYGGKACVG 535
Db 330 -----PCP-DG-----YTGNGITCS----- 343

Qy 536 DVQERQCNKRSCPVDGCLSNPCFPGAQCSFPDGSWCGSPGVFLG----- 583
Db 344 -----DVDECKYHPCYGVRCNTNLAPG-FRCDACPVGFTGMVGVGINFAK 389

Qy 584 -NGTHCEDLDECALVPDLCFSTSKVPRCVNTQPGPHCLPCPPRYRGNQPVGVGLEAAKTE 642
Db 390 TNKQVCTDVEDECR---NGACVLNS---ICINTLGSRYRCPCPKPGYTGDTQTRG----- 435

Qy 643 KOVCEPENPKDKTHN-CHKHAECIYLGHFSDPMYKCECQTGYAG-DGLICGEDSLDGGW 700
Db 436 ---CFTERSCRNPEQNPCSVHAQCI---EERQGDVTCVGVGVWAGRAGYVCGKDVIDSY 489

Qy 701 PNLNLVCAATNATYHCIKNCPLPNSGODEFDKDGIGDACDDDDNDGVDTEKONCQLLF 760
Db 490 PDELPCCSAR---NCKKDNCKYVPSNGDEADRDGIGDACDADGADGDLNEQDNCVLTH 546

Qy 761 NPROADYDKDEVGRCDCNCPYVHNPAQIDTDNNGSGDACSVDIDGDDVDVFNERNDCNPPYVN 820
Db 547 NVQQRNTDKDIFGDACDNCRGVNLNNDQKDTDGDGKGACDADDMDGDKNILLNDCNPRVFN 606

| | | | |
|----|------|---|------|
| QY | 821 | TQORTDSDGVGDHCDNCPLVHNPQTDVNDLVGQCDNNEIDDDGHQNNQDNCPTIS | 880 |
| Db | 607 | RQQRDRDSDGVGDACDSCPDVSNPNQSDVDNDLVGSDCDTNQDSGDGHQDSTDCNPTVI | 666 |
| QY | 881 | NANQADHRRDGGDADCPDDDDNDGVDP-----DRDNCRLVFNPDQEDLDGDRGDIKDDF | 936 |
| Db | 667 | NSAQLDTOKDGIQDECDDDDDNDGMPDLFPFGPDNCRCLVNPNAQEDSNNDGVGDICEADF | 726 |
| QY | 937 | DNDNTPDIDDVCPENNAISETDFRNFOMVPLDPKGTQIDPNWVIRHOGKELVQTANSDP | 996 |
| Db | 727 | DQKVIDRIDVCPENAEITLDFRAYQTVVLDPEGDAQIDPNWVILNQMEIVQTMNSDP | 786 |
| QY | 997 | GIAVGDFEFGYDFSGTFYVNTDRDDYAGFVFGYQSSRFYVVMKQVQTYWEDQPTR | 1056 |
| Db | 787 | GLAVGYTAFNGVDFEGTFHVNTQDDDYAGFIFGYQSSRFYVVMKQTEQTYWQATPR | 846 |
| QY | 1057 | AYGYSVSLKVYNTGTGCEHLRNALWHTGNTPGQVRLWHDPRNIGWKDYTAYRWHLTH | 1116 |
| Db | 847 | AVAEPGIQLKAVKSKTGPEHLRNSLWHTGDTSDQVRLWKDSRNVGWKDKVSYRWFLQH | 906 |
| QY | 1117 | RPKTGYIRVLVHEGQVMADSGPIYDQTVAGGRLGLVFQSEMYVFDLKYECD | 1171 |
| Db | 907 | RPOVGYIRVRYEGSELVADSGVTIDTMRGGRGLGVFCFQSENIWISNLKYRCND | 961 |

Search completed: August 9, 2002, 10:06:34
Job time: 294 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:01:19 ; Search time 46.31 Seconds
(without alignments)
4378.105 Million cell updates/sec

Title: US-09-822-682-2
Perfect score: 6605
Sequence: 1 MWRLVILALVWPFSTQAGH.....FVFSQEMVYFSLKYECRDI 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_19.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mnc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertibrate.*
- 14: sp.unclassified.*
- 15: sp.rvirus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 1988 | 30.1 | 817 | 4 | O14592 |
| 2 | 1966 | 29.8 | 755 | 11 | O9R0G6 |
| 3 | 1959 | 29.7 | 755 | 6 | O9B880 |
| 4 | 1913.5 | 29.0 | 851 | 13 | O42507 |
| 5 | 1902 | 28.8 | 863 | 11 | O9QYS3 |
| 6 | 1901 | 28.8 | 963 | 11 | O921T2 |
| 7 | 1821 | 27.6 | 1060 | 5 | O968S4 |
| 8 | 1805 | 27.3 | 1024 | 5 | O9VM97 |
| 9 | 962 | 14.6 | 287 | 13 | O91002 |
| 10 | 833 | 12.6 | 181 | 13 | P79786 |
| 11 | 639.5 | 9.7 | 376 | 4 | O15667 |
| 12 | 580 | 8.8 | 128 | 6 | O28195 |
| 13 | 505 | 7.6 | 5636 | 4 | O96RW7 |
| 14 | 504 | 7.6 | 2673 | 4 | O96SC3 |
| 15 | 442 | 6.7 | 585 | 6 | O9TTS5 |
| 16 | 429 | 6.5 | 1444 | 5 | Q17591 |

| | | | | | |
|----|-------|-----|------|----|--------|
| 17 | 418.5 | 6.3 | 4123 | 4 | O75851 |
| 18 | 417.5 | 6.3 | 1637 | 6 | O9XSV8 |
| 19 | 410 | 6.2 | 2327 | 13 | O9IBG7 |
| 20 | 405 | 6.1 | 1062 | 5 | O19204 |
| 21 | 400 | 6.1 | 1202 | 4 | O9P283 |
| 22 | 399 | 6.0 | 788 | 5 | O22631 |
| 23 | 390 | 5.9 | 89 | 11 | O9R150 |
| 24 | 381 | 5.8 | 229 | 6 | O28194 |
| 25 | 379 | 5.7 | 1121 | 5 | O966P9 |
| 26 | 379 | 5.7 | 1148 | 5 | O966Q0 |
| 27 | 378.5 | 5.7 | 1081 | 5 | O9U631 |
| 28 | 377 | 5.7 | 103 | 6 | O9N2C3 |
| 29 | 376.5 | 5.7 | 1083 | 5 | O9VTT0 |
| 30 | 365.5 | 5.5 | 815 | 4 | O96JS2 |
| 31 | 358.5 | 5.4 | 425 | 6 | O02661 |
| 32 | 341 | 5.2 | 2189 | 5 | O9BI05 |
| 33 | 341 | 5.2 | 2524 | 5 | O9GPA5 |
| 34 | 333.5 | 5.0 | 807 | 6 | O9GLX9 |
| 35 | 332.5 | 5.0 | 624 | 4 | O94862 |
| 36 | 331.5 | 5.0 | 807 | 4 | O9HCB6 |
| 37 | 322 | 4.9 | 802 | 13 | O9W770 |
| 38 | 322 | 4.9 | 2471 | 11 | O9QW30 |
| 39 | 321 | 4.9 | 2470 | 11 | O35516 |
| 40 | 319.5 | 4.8 | 769 | 5 | O00816 |
| 41 | 318.5 | 4.8 | 2531 | 5 | O16004 |
| 42 | 317 | 4.8 | 1360 | 5 | O9TYK4 |
| 43 | 312.5 | 4.7 | 861 | 11 | O9QW58 |
| 44 | 312 | 4.7 | 2471 | 4 | O04721 |
| 45 | 311.5 | 4.7 | 257 | 5 | O966K7 |

ALIGNMENTS

RESULT 1
O14592 PRELIMINARY; PRT; 817 AA.
AC O14592;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COMP_HUMAN.
GN COMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Gordon L., Kyle A., Ramirez M., Stillwagen S., Garnes J., Panganan L.,
RA Christensen M., Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A.,
RA Olsen A.O., Carrano A.V.;
RT "Sequence analysis of an -1 Mb region containing the MEF2B gene in
RT 19p12.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC003107; AAB86501.1; -;
DR HSSP; P35444; 1VDF.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 11.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 817 AA; 89148 MW; 70551980BE221855 CRC64;

Query Match 30.1%; Score 1988; DB 4; Length 817;

| | | | |
|--|--|--|------|
| Best Local Similarity 54.1%; Pred. No. 1.2e-148; | | Matches 361; Conservative 73; Mismatches 175; Indels 58; Gaps 6; | |
| QY | 549 | PVDCGLSNPCFPGAQCSFPDGSWCSGCPVGLGNGTHCEDLDECALVPDIDCFSTSKVP | 608 |
| Db | 87 | PLLCAPGCFPCGACVACIQTESGA-RCGPCAGTGTGSHCTDVNECNAP--CFPR--V | 140 |
| QY | 609 | RCVNTOPGFHCLPCPPRYRGQVGVLENAKTEKQVCEPENCKDKTHN----- | 658 |
| Db | 141 | RCINTSPGFCEACPFGYSGTHQGVGLAFKANKQVCVDINECETGQHCNCPVNSVCINT | 200 |
| QY | 659 | -----CHKHAECIYLGHSFSDPMYKCECOT | 682 |
| Db | 201 | RGSFQCGPCQPGFVGQASCCORRAQRCFCPDGSPSECHHADCVL---ERDGRSVCVAV | 257 |
| QY | 683 | GYAGDGLICGEDSLDGPWNLNLVCAATNATYHCIKDNCPLHNSGQEDFKDGIACDD | 742 |
| Db | 258 | GWAGNGILCGRDLDLGGFDEKLCRCER---QCRKDCNCTVPNSGQEDVDKIGDADCP | 314 |
| QY | 743 | DDNDGVTDEKNCQLLFPNROADYDKDEVDRCDCNCPVHNPADTDTONNGEGDACSVD | 802 |
| Db | 315 | DADGDGVFNKNCPLVRNPQDRTDEKWDGACDNCRSQKNDQKDTDDGGRGACDD | 374 |
| QY | 803 | IDGDDVFNERNDCPYVNTDQRTDGDGVDHCDNCPVHNPQTDVDNDLVGDQDQDNE | 862 |
| Db | 375 | IDGDIRNQADNCPVNSDKSDSDGIGDADCNCPQKSNPDQADVDHDFVGDACDSQ | 434 |
| QY | 863 | DIDDDGHQNNQDCPYISNANQADHDDRGQDADCPDDNDGVDPPDRDNCRLVFNPDQED | 922 |
| Db | 435 | DQDGDGHQSRDCNCPVNSAQSDSDHGGQDADCDDDNDGVDPSRDCNRLVFNPDQED | 994 |
| QY | 923 | LDGDDGRDICKDQDFNDNDPDIIDVCPENNAISETDFRQFQVPLDPKGTQTDPNWVIR | 982 |
| Db | 495 | ADRDGVDGVDQDDFKVVDKIDVCPENAEVTLTDFRAFQTVLLDPEGDAQIDPNWVIL | 554 |
| QY | 983 | HQKELVQTANSPDGIAGVDFEGSVDGTFYVNTDRDDYAGVFGYQSSSRFYVVMW | 1042 |
| Db | 555 | NOGREIVQTMNSDPLAVGTAENGVDGTFYVNTDQDQYAGVFGYQSSSRFYVVMW | 614 |
| QY | 1043 | KQVQTYWEDQPRAYGYSVLSKVYNTGTGTEHLRNALWHTGTPGQVRLWHPDPRNI | 1102 |
| Db | 615 | KQVQTYWEDQPRAYGYSVLSKVYNTGTGTEHLRNALWHTGTPGQVRLWHPDPRNV | 674 |
| QY | 1103 | GWKDYAYRWHLTHRPKTYGIRVLYHEGQVADSGPIYDQTYAGRGGLGFVFSQEMWYF | 1162 |
| Db | 675 | GWKDYAYRWHLTHRPKTYGIRVLYHEGQVADSGPIYDQTYAGRGGLGFVFSQEMWYF | 734 |
| QY | 1163 | SDLKYEC | 1169 |
| Db | 735 | ANLYRC | 741 |
| RESULT 2 | | | |
| QY | 9066 | PRELIMINARY; PRT; 755 AA. | |
| AC | Q9066 | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | |
| DT | 01-DEC-2001 (TrEMBLrel. 19, Last annotation update) | | |
| DE | CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR. | | |
| GN | COMP. | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| OX | NCBI_Taxid=10090; | | |
| RP | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=CARTILAGE; | | |
| RA | Fang C., Carlson C.S.; Leslie M.P., Tulli H., Stolerman E., Ferris R., | | |
| RA | Ni L., Di Cesare P.E.; | | |
| RT | "Molecular Cloning, Sequencing, Tissue and Developmental Expression of | | |
| RT | Mouse Cartilage Oligomeric Matrix Protein (COMP)."; | | |
| RL | J. Orthop. Res. 0:0-0(1999). | | |

| | | | |
|--|---|--|------|
| DR | EMBL; AF033530; AAD01972.1; -- | | |
| DR | HSP; F35444; 1VDF. | | |
| DR | MGI; MGI:88469; Comp. | | |
| DR | InterPro; IPR002048; EF-hand. | | |
| DR | InterPro; IPR000561; EGF-like. | | |
| DR | InterPro; IPR001881; EGF_Ca. | | |
| DR | InterPro; IPR003367; tsp_3. | | |
| DR | Pfam; PF02412; tsp_3; 9_3. | | |
| DR | SMART; SM00179; EGF_Ca; 2. | | |
| DR | SMART; SM00001; EGF_like; 2. | | |
| DR | PROSITE; PS00018; EF_HAND; UNKNOWN_1. | | |
| DR | PROSITE; PS01186; EGF_2; 1. | | |
| DR | PROSITE; PS01187; EGF_Ca; 2. | | |
| KW | Calcium-binding; EGF-like domain; Glycoprotein; Matrix protein; | | |
| KW | Repeat; Signal. | | |
| FT | SIGNAL 1 19 POTENTIAL. | | |
| FT | CHAIN 20 755 CARTILAGE OLIGOMERIC MATRIX PROTEIN. | | |
| SEQ | SEQUENCE 755 AA; 82352 MW; 7DDCF443589A0B7 CRC64; | | |
| Query Match 29.8%; Score 1966; DB 11; Length 755; | | | |
| Best Local Similarity 54.4%; Pred. No. 6.1e-147; | | | |
| Matches 362; Conservative 72; Mismatches 180; Indels 52; Gaps 7; | | | |
| QY | 549 | PVDCGLSNPCFPGAQCSFPDGSWCSGCPVGLGNGTHCEDLDECALVPDIDCFSTSKVP | 608 |
| Db | 85 | PVPLCAPGSCFQVVCSETATGA-RCGPCPPGTGTGSHCTDVNECNAP--CFPR--V | 138 |
| QY | 609 | RCVNTOPGFHCLPCPPRYRGQVGVLENAKTEKQVCEPENCKDKTHCHKHAECI-- | 666 |
| Db | 139 | RCINTSPGFCEACPFGSGTHQGVGLTFAKSNKQVCVDINECETGQHCNCPVNSVCNT | 198 |
| QY | 667 | -----YLG-----HF-----SDPMYKCECOTGA | 695 |
| Db | 199 | RGSFQCGPCQPGFVGQDTSQRRGQHFCDGSPSCHEKANCVLERDSRSCVCAVGA | 258 |
| QY | 686 | GDGLICGEDSLDGPWNLNLVCAATNATYHCIKDNCPLHNSGQEDFKDGIACDDDD | 745 |
| Db | 259 | GNGLLCGRDLDLGGFDEKLCRCER---QCRKDCNCTVPNSGQEDVDKIGDADCPAD | 315 |
| QY | 746 | NGVTEKNCQLLFPNROADYDKDEVDRCDCNCPVHNPADTDTONNGEGDACSVDIG | 805 |
| Db | 316 | GDGVPNEQDCNCPVLRNPQDQNSDSDGVDGACDNCRCRKNDDQKDTDLGGRGACDD | 375 |
| QY | 806 | DDVFNERNDCPYVNTDQRTDGDGVDHCDNCPVHNPQTDVDNDLVGDQDQDNEID | 865 |
| Db | 376 | DRIRNVADNCPVNPFDQSDSDGVDGACDNCPCQKDNPDQDQDQDQDQDQDQDQD | 435 |
| QY | 866 | DDGHNQNDNCPYISNANQADHDDRGQDADCPDDNDGVDPPDRDNCRLVFNPDQEDLDG | 925 |
| Db | 436 | GDGQDSRDCNCPVNSAQDSDHDKGKACDNDGVDPSRDCNRLVFNPDQEDNDR | 495 |
| QY | 926 | DGRGDIKDDDFNDNDPDIIDVCPENNAISETDFRQFQVPLDPKGTQTDPNWVIRHOG | 985 |
| Db | 496 | DGVGDACQDGFADKVIDKIDVCPENAEVTLTDFRAFQTVLLDPEGDAQIDPNWVILNQ | 555 |
| QY | 986 | KELVQTSNDSPGIAGVDFEGSVDGTFYVNTDRDDYAGVFGYQSSSRFYVVMWQV | 1045 |
| Db | 556 | MEIVQTMNSDPLAVGTAENGVDGTFYVNTDQDQYAGVFGYQSSSRFYVVMWQV | 615 |
| QY | 1046 | TQTYWEDQPRAYGYSVLSKVYNTGTGTEHLRNALWHTGTPGQVRLWHPDPRIGWK | 1105 |
| Db | 616 | EQTYQANPFRVAVPEGIQKAVKSTGPGQELRNALWHTGTPASQVRLWHPDPRVGVK | 675 |
| QY | 1106 | DYTAIRWHLTHRPKTYGIRVLYHEGQVADSGPIYDQTYAGRGGLGFVFSQEMWYFSDL | 1165 |
| Db | 676 | DKTSYRWFLQHRPQVGIIRVRYEGPELVADSNVILDTAMRGRLGVFCFSQENIIWAL | 735 |
| QY | 1166 | KYECRD | 1171 |
| Db | 736 | RYRCND | 741 |

| | | | |
|---|--|--|------|
| Db | 553 | NOGMEIVQTNMSDPLAVGYTAFGVDEGTEGTHVNTVDDYAGTIFGYQDSSEFYVVM | 612 |
| Qy | 1043 | KQVTOIWEQDPFRAYGYSVLSKVNSTGTGHELRNALWHTGTPGOVTRLWHDPRI | 1102 |
| Db | 613 | KMEQTYWQANPRVAVAEPIQLKAYKSTGEGEQLRNALWHTGTAQVRLLMKDPRI | 672 |
| Qy | 1103 | GWKDYTAIRWHLTHRPKTGYIRVLVHEGKOVNADSGPIYDQYAGRGLGTFVSEMYVF | 1162 |
| Db | 673 | GWKDKTSYRWFQHRROVGYIRVRFEGPELVADSNVLYLDTMRGRLGVCFSENIIV | 732 |
| Qy | 1163 | SDLKYECRD 1171 | |
| Db | 733 | ANLRYCRND 741 | |
| RESULT 4 | | | |
| Qy | 042507 | PRELIMINARY; PRT; 851 AA. | |
| Id | 042507 | | |
| OC | A042507 | | |
| DT | 01-JAN-1998 (TrEMBLrel. 05, Created) | | |
| DT | 01-JAN-1998 (TrEMBLrel. 05, Last sequence update) | | |
| DT | 01-DEC-2001 (TrEMBLrel. 19, Last annotation update) | | |
| DE | THROMBOSPONDIN-3 (FRAGMENT). | | |
| OS | Xenopus laevis (African clawed frog). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; | | |
| OC | Xenopodinae; Xenopus. | | |
| OX | NCBI_TaxID=8335; | | |
| ON | [1] | | |
| RA | SEQUENCE FROM N.A. | | |
| RP | urry L.A., Whittaker C.A., Duquette M., Lawler J., Desimone D.W.; | | |
| RT | "Expression of thrombospondin-1, thrombospondin-3 and thrombospondin-4 | | |
| RL | mRNAs during Xenopus laevis embryogenesis."; | | |
| RL | Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL: U92805; AAB69765.1; -. | | |
| DR | HSSP; P35555; 1ENN. | | |
| DR | InterPro: IPR000561; EGF-like. | | |
| DR | InterPro: IPR001881; EGF_Ca. | | |
| DR | InterPro: IPR003367; tsp_3. | | |
| DR | Pfam: PF02412; tsp_3; 8. | | |
| DR | SMART: SM00179; EGF_CA; 2. | | |
| DR | SMART: SM00001; EGF_like; 2. | | |
| DR | PROSITE: PS01186; EGF_2; 1. | | |
| DR | PROSITE: PS01187; EGF_CA; 2. | | |
| DR | Calcium-binding; EGF-like domain; Glycoprotein; Repeat. | | |
| FW | NON_TER | | |
| FT | 1 | | |
| SEQUENCE | 851 AA; 93307 MW; 23A031EB8DA49E9 CRC64; | | |
| Query Match 29.0%; Score 1913.5; DB 13; Length 851; | | | |
| Best Local Similarity 51.7%; Pred. No.1.le-142; | | | |
| Matches 356; Conservative 78; Mismatches 189; Indels 65; Gaps | | | |
| Qy | 537 | VBERQMC---NRRSCPVDGCLSNPCFPGAQC--SSPPDGWSWCGSCPVGFLNGTHCEDLD | 592 |
| Db | 158 | IMECQVCGFHEHS---RCNPNPCFPGVDCMETEYEGYRCGCPCTGSGNGTYCADID | 213 |
| Qy | 593 | ECALVPDICFSTSKVPRCVNTQPGHCLPCPPRYRGNOVPVGLAETKQVCEPENPC | 652 |
| Db | 214 | EGSHA-NPCFSGSK---CINTSPGRCEHCPRGYKGNVTSGVGVDIYAKASKQVCRDIDEC | 269 |
| Qy | 653 | KKTHN-----CHKHAECIV | 667 |
| Db | 270 | NDGNGGCDANSICNTNTVGSFRCGPKPGFVGNQTVGCVPKKSCASPAFNPCHVNGHCVF | 329 |
| Qy | 668 | LGHFDPMYKCECOTQYAGDGLICGEDSDLDGWPMLNLVLCATNATYHCIKDCNCPHLPSG | 727 |
| Db | 330 | ERNGD---ITCACNVGWAGNGYTCGRDITLDGYDEPMPICDN-NKHCQDNCRCSTNSG | 385 |
| Qy | 728 | QEDFKDGTGACDDDDNDNDGVTDEKDNQCLLFNRPQADYDKDEVGRCDCNCPVHNPAQ | 787 |
| Db | 386 | QEDANDGTGDCDEADQDGTGKNVEDNCRVLPNKDQONSOTDSFGDACDCNCPVNNQD | 445 |

```
QY 788 IDTDNNGEGDACSVDIDGDDVFNERNDCPYVINTDQRTDGDGVDGHCNCPVHNPDQT 847
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 KDTDANGESDADNDIDGDDGTPNMLDNCYPNPLQTDRLDGVGDACDSCPEASNPQA 505
QY 848 DVDNDLVGQCCONNEDIDDDGHQNNQDNCPIYSNANQADHRDGGDGDPCDDNDGVPD 907
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 DADSLVAGDMCDTQNRDGDGHQDTKNCPCDIPNSSLSDSDNDGKGDGCDQDDNDGIPD 565
QY 908 ----DRDNCRLVFNPDQEDLDGDDGDKDDFNDNDPIDDVCPENNAISETDFRNFQ 963
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 YIPGPDNCRILHNPQKSDGDDGVGDVCEDEFDNDTVIDPLDVCPEAEVTLTDFRAYQ 625
QY 964 MVLDPKGTQIDPNMWIRHOGKELVQANSDPGIAVGFDFSGYSDFSTGYVNTDRDD 1023
Db 626 TVILDPEGDAQIDPNWVNLNGMEIVQTMNSDPLGLAVGYTAFNGVDFFGTFHVNTVTD 685
QY 1024 YAGFFVGQSSSRFVVMWQVQTYWEDQTPRAYGSGVSLKVNNSTTGTGHEHLRNALW 1083
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 YAGFFYSQDSSSFVVMWQTEQTYWQATPFRAVAEPGLQLKAVKSTGPGELRNALW 745
QY 1084 HTGNTPGVRLTWHDPNIGWKDYATYRWHLTHRPKTYIRVLVHEGQVMADSGPIYDQ 1143
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 746 NTHQTDQVRLNWKDPRNVGKDKTSYRWQLMHRPQVGYIRVLYEGVDLVADSGVLI 805
QY 1144 TYAGRLGLFVSQEMYYFSLKYECD 1171
Db 806 TMRGRLGVFCFSQENLIWSNLOYRCD 833

RESULT 5
ID Q90YS3 PRELIMINARY; PRT; 863 AA.
AC Q90YS3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE THROMBOSPONDIN 4 (FRAGMENT).
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AAD32714.1; -.
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Thbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
FT NON_TER
SQ SEQUENCE 863 AA; 95339 MW; 68E3EE6846728E4D CRC64;
```

Query Match 28.88; Score 1902; DB 11; Length 863;
Best Local Similarity 53.56; Pred. No. 8.7e-142;

```
Matches 358; Conservative 71; Mismatches 182; Indels 58; Gaps 10;
QY 549 PVDCLSNPCFPQAQCSFFPDGSSNSGSCPCVCFGLNGTHCEDLDCALVPDICEFSTKVP 608
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 PTHCHDSSPCFGRVCTDTRDG-FQCPCPDGYTGNGITCSVDVECKYHP--CYPG---V 241
QY 609 RCVNTQPGFHLCPPPRYRGNQPVGVGLEAAKTEKQVC-----EPEN----- 650
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 RCVNLAPGFRCDACPVGFTGPMVGVGINFAKTNKQVCTDVEQCGACVLNSICINTLG 301
QY 651 -----PK-----DKTHNCHK-----HACIYLGHFSDPMYCECOTGAG 686
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 SYRGPCKPGYTGQTRGCKTERSCRNPEQNPCSVHAQCI---EERQGDVTCVCGWAG 358
QY 687 DGLICGEDSLDGNPLNLVLCATNATYHCIDKNCPHLPNSGOEDDKGIGDACDDDDN 746
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 DGYVCGKDVDDISVPDEELPCSA---NCKDKNCKYVNSQEDADRIGIGDGEDADG 415
QY 747 DGVTEKDNKCOLLNPQADYDKDEVDGDCNCPVYHNPQAQIDTDNNGEGDACSVDIDG 806
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 DGIILNEQNCVLTHNIDQNSDKDIFGDACDNCRMVLNNDKQDTDGDCRGDADDDMDGD 475
QY 807 DVFNERNCPVYNTDQRTDGDGVDGHCNCPVHNPDQTDVNDLVGDCCDNHEDIDD 866
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 GIKNILDNCPRVNRDQDQDDVDGADCDSCPDVSNPNQSDVDNDLVGDSCDTNQDSDG 535
QY 867 DGHQNNQDNCPIYSNANQADHRDGGQGDACDPCDDNDNGVDP-----DRNCRLVFNPDQED 922
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 DGHQDSTDNCPTVINSSQLDQDKGIGDECDDDDDNDGIDPLVPPGPNCLVPPAQED 595
QY 923 LDGDRGDIKDDFDNDNIPDIDVCPENNAISETDFRNFQVPLDPKGTQIDPNWVIR 982
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 SNNDGVGDICEADFQDQVIDHIDVCPENAEITLTDFRAYQTVVLDPEDGAQIDPNWVVL 655
QY 983 HOGKELVQANSDPGIAVGFDFSGYSDFYVNTDRDDYAGVFCYQSSSRFVVMW 1042
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 NQGMIEVQTMNSDPLGLAVGYTAFNGVDFEGTFHNTQDQDQDQDQDQDQDQDQDQ 715
QY 1043 QVTTQTYWEDQTPRAYGSGVSLKVNNSTTGTGHEHLRNALMHTGNTPGQVRLTWHDP 1102
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 716 KQTEQTYWQATPFRAVAEPGLKAVKSKTGPGEHLRNLWHTGTDSDQVRLWKDSRV 775
QY 1103 GWKDYATYRWHLTHRPKTYIRVLVHEGQVMADSGPIYDQTYAGRLGLFVSQEMYYF 1162
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 776 GWKDKSVYRWFLQHRPQVGYIRVRYEGSELVADSGVTIDTMRGRLGVFCFSQENI 835
QY 1163 SLDKVECD 1171
Db 836 SNLKYRCND 844

RESULT 6
ID Q921T2 PRELIMINARY; PRT; 963 AA.
AC Q921T2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE THROMBOSPONDIN-4.
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF102887; AAC73003.1; -.
DR HSSP; P35444; 1VDF.
```


Db 531 VCDRAVCLMDNIRHKCHCNVWAGNGLICGRDITVDGWPQAIGC---PELHCQRDNC 587
 Qy 721 PHLPNGSGEDFKGIGIGACDDDDNDGVDTEDEKNCQLLFPNRQADYKDEVDGRCDNCP 780
 Db 588 PKLPSNGSGEDADLGGHGGCGDADGNDVQNSQDNCWLAYNTEQLDSDGDKYGVDCDNCV 647
 Qy 781 YVNEPAQIDTNNNGEGDACSVDIDGDDVFNERNDCNPPYVNTDQRTDGDGVDGHDGDCNCP 840
 Db 648 LKYNPRLQDITDEGLGDECDGIDNDSTPNALDNCPLPNSQSDVDNDGVDGADCDNCPN 707
 Qy 841 VHNPDQTDVNDLVDGQDNNEDIDDDGHQNNQDNCPIYSNANQADHRODGGDCACTDPPD 900
 Db 708 LFNPKQKQDMDFVGDACHROIDGDDGVPSNLDNCPMVNSDQLDITDGDGTGDCDDDM 767
 Qy 901 DNDGVPDRDNCRLVFNPDQEDLDGDRGDCIDKDFDNDNIPDIDVCPENNAISETDFR 960
 Db 768 DGDGIPNKNKCPALAKNPKDFNRNGKDCEDDEDVGVPMGMDNCPNNSMIHHTDFR 827
 Qy 961 NFQMPVLPQKGTQIDPNVIRHOGKELVQTANSDFGIAVGFDFSGVDFSGTFYVNTDR 1020
 Db 828 TLQITPLDPLKGLSQADPNWVYHANGTEIVQTLNSDPLGLAVGKDAFGVDGDFGTFFINDT 887
 Qy 1021 DDYAGFVFGYSSRFYVVMKQYQTYWEDQPTRAYGSGVSLKVVNSTTGTGTEHLRN 1080
 Db 888 DDYAGFVFGYSSRFYVVMKQYQTYWEDQPTRAYGSGVSLKVVNSTTGTGTEHLRN 947
 Qy 1081 ALWHGNTPGQVRLTWHDPNRNIGWDTYAYRPHLTHRPKTYGIRVILVHVGQVMDASGPI 1140
 Db 948 SLWHEGNTDGEARLQKPKNTAWKERTSYRSLVHRPFAIGLIRLOMHEGRLIFDSGNV 1007
 Qy 1141 YQQTAGGRLGFLVFSQEMVYFSLKYECRD 1171
 Db 1008 FDSLTKGRLGVCFPSQRMILWNSNLOYKCNN 1038
 RESULT 8
 Q9VM97 PRELIMINARY; PRT: 1024 AA.
 AC Q9VM97
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CG11326 PROTEIN.
 GS CG11326
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Chertis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003614; AAF52425.1; -.
 DR HSSP: P35555; 1EMN.
 DR FlyBase: FBgn0031850; Tsp.
 DR InterPro: IPR000152; Asx_Hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR003367; tsp_3.
 DR Pfam: PF00008; EGF_3.
 DR Pfam: PF02412; tsp_3; 8.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF_like; 3.
 DR PROSITE: PS001010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 1024 AA; 114398 MW; 88F279D6804610AD CRC64;
 Query Match 27.3%; Score 1805; DB 5; Length 1024;
 Match Local Similarity 34.0%; Pred. No. 5, 4e-134;
 Matches 407; Conservative 134; Mismatches 354; Indels 302; Gaps 31;
 Qy 77 IMRQKEGFFLTAQLKQDGKSGRTLLALGPGLSQ-RQFEIVSNGPADLDTLYWDGTRH 135
 Db 6 MLKHKMSFFL-----DRKQQRVTLDISANGATESRNFEPININETSI----- 48
 Qy 136 VVSLVDVGLADSQWKNVTVOAGETSLHYGCDLLDSFALDEPFYEHQAQSRMYAVK 195
 Db 49 -----RSLALQFSKNRITLHVDCAKASTHDDID-----MNLAKLYTQMD 86
 Qy 196 SARESHFRGLLQNVHLVFNSEVEDILSKGCGQOG---QGAELNAISENTEITRLGPHV-- 250
 Db 87 DPVIKLFRRKYPLH--FDGDMHSLQRANCKQGNHRGRNRLNKLITERGELFPRILG 144
 Qy 251 ----TTEYV-----GPSSRRPEV---CER----- 268
 Db 145 YLOPTMYIYSAWYAPIFAEKNKKRDVRGWEPTIAREGVVDHRHQEVPTDVERGDIPLV 204
 Qy 269 --SCEELNMVQELSLHLVYNQPSLENLKRYSNDNQFLWELIGGPKTRNMSACWDGRF 326
 Db 205 NGDCEDA--LARSLSLLALVRLREDVAHQEITAYL-----RM 242
 Qy 327 FAENETWVDSCTTCT-CKKFKTICHQITC-PPATCASPSFVEGCC-----PSCLHSVD 379
 Db 243 LLEN-----CAGCKNPLTDTNQLRIEPCDCRSANPCYPGVCELDASAAGPCRGH--- 289
 Qy 380 GEEGSPWAEWTQCSVTGSGTQQRGRSCDVSTNTCLGPSIQTRACSLSKDTRTRDQGG 439
 Db 290 -----CPLGFIGDGKSK-----PGV---TCAHMKVPYGVQ----- 317
 Qy 440 WSHWSPWSSCSVTCGVGNITRILCNPSVPQMGKNGKSGRETRKACOGACPIDRWSP 499
 Db 318 -----CHDTV-----NGAQCD----- 328
 Qy 500 WSPWSACTVTCAGGIRTRVNCSPSPQYGGKACVGDVQEROMCNKRSCPDGCLSNPCF 559


```
Db 329 -----SCPAGYEGDGTSLRNP-----CLDTPCP 353
QY 560 PQAQC---SSFPDGSWCGSPVGLNGTHCEDLDECALVPDICTFSTSKVPRCVNTQPGF 617
Db 354 SGAQCLQGVGPP-YFHCISCPMGHEVNGTSCROMNEC-LLYDPC---DELATCTNLSPGF 408
QY 618 HCLPCPPRYGNQPVGVGLE--AAKTEKQVCEPENCKDKTTHCKHAECI-----666
Db 409 QCSPCVGFDDGTHAGFYADYTSVYRRQTCLDVDECRGFFRCPHEHSTCINEIGSYRQO 468
QY 667 -YLGHSFDPWY-----KCECQTGYAGDGLICGD 694
Db 469 CHEGYVNTGYTSCLDSSVFMCPDGTVCDRNAVCLRMNIRKHCNVCNAGNGLICGRD 528
QY 695 SLDJGWPNLNLVCATNATYHCKIDNCPLHNSQGEFDKDGIGDADCDDDDDNDGVTDKRD 754
Db 529 TDVDGWPDAIGC---PELRCQRDNCPLKPLNSQGEADLDGHDGDCDDADGDNVQNSQD 585
QY 755 NCOLLENPRQADYKDEVDGDRDNCPCYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNR 814
Db 586 NCWLAYNTQLSDSGDKVGDCVDCNCLYNPRQLDDEGLGDECDGIDNDSIPNALDN 645
QY 815 CPVYVNTDQRTDGGVGDHDCNCPVLVHNPQDQTVNDLVGDCQDNNEDIDDDGHONNOD 874
Db 646 CPLLPNPSQSDVNDGVDGACDNCPLPNPDQKDRDMDFVGACHRDIDGDDGVNPSLD 705
QY 875 NCPYISNANOADHRDQGDGACDPPDDNDGVPDDRNCRLVFNPDQEDLDGGRGDIKCD 934
Db 706 NCPMVSNSQLDGTGDTGDECDDDMDGDIIPNYKDNCPKAKNPKQDDFNKNGKGSCE 765
QY 935 DFDNDNIPDIDVCPENNAISETDFRNFQWPLDPKGTQIDPNVIRQKGLVQTANS 994
Db 766 DEDVDGVPNGMDNCPNNSIHHTFTLTQIDPLPKLSQADPNVYVHANGTEIVOTLNS 825
QY 995 DPGIAGVFDFGFSVDSFGTFYVNTDRDDYAGFVFGYQSSRFYVYVWKKQVQTVMEDQP 1054
Db 826 DPLAVGKDAFGVDFGTFYINDTDDYAGFVFSYQSSYKYVYVWKKGTQTYWEP 885
QY 1055 TRAYGSGVSLKVNSTTGTGHEHLNWLHNTGTPQVTLHDPNIGKWKDYATRWHL 1114
Db 886 FTASAPGIIQIKLVNTEGPGPMRNSLWHEGNTDGEARLLWKDKPKNAWKERTSYRSL 945
QY 1115 THRPKTYIRLVLCQVMAQSGPIYDQTYAGRGGLGFVFSQEMVYFSDLYCECRD 1171
Db 946 VHRPAIGLRLQMHENRILFDSGNVDFSTLKGRLGVCFSGORMIWSNLQIKCNN 1002

RESULT 9
ID Q91002 PRELIMINARY; PRT; 287 AA.
AC Q91002;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THROMBOSPONDIN-4 (FRAGMENT).
GN Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96016561; PubMed=7496039;
RA Tucker R.P., Adams J.C., Lawler J.;
RT "Thrombospondin-4 is expressed by early osteogenic tissues in the
chick embryo."
RL Dev. Dyn. 203:477-490(1995).
DR EMBL; L27263; AAA9960.1; -
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 8.
FT NON_TER 1
```

```
FT NON_TER 287 287
SQ SEQUENCE 287 AA; 30880 MW; 9B3B0E413C95FC3F CRC64;

Query Match 14.6%; Score 962; DB 13; Length 287;
Best Local Similarity 61.6%; Pred. No. 5.1e-68;
Matches 175; Conservative 31; Mismatches 74; Indels 4; Gaps 1;

QY 761 NFRQADYKDEVDGDRDNCPCYVHNPAQIDTDNNGEGDACSVDIDGDDVFNERNRNCYVYN 820
Db 4 MYNQNSDEDIFEGDADNCRSVLNNDQRTDGDGDKGACDDDDMDGDKIKLLDNCORIPN 63
QY 821 TDQRTDGGVGDHDCNCPVLVHNPQDQTVNDLVGDCQDNNEDIDDDGHONNODNCPYIS 880
Db 64 QDQEDKNDGVDGACDSCPTVSNPNQSDNDLVGDCSTNQDSDGDDGDDSTDNCPITII 123
QY 881 NANOADHRDQGDGACDPPDDNDGVPD----DRDNCRLVFNPDQEDLDGGRGDIKDDF 936
Db 124 NSSQLDTKDGLGDECDDEDDDGIPLDLPDNCRLVFNPDQEDDNGDGVDCESDF 183
QY 937 DNDNIPDIDVCPENNAISETDFRNFQWPLDPKGTQIDPNVIRQKGLVQTANSOP 996
Db 184 DQDTVIDRIDVCPENAEITLDFRAYQTVYLDPEGAQIDPNVYVNLQGMIEIVQTINSOP 243
QY 997 GIAVGDFEGSVDFSGTFYVNTDRDDYAGFVFGYQSSRFYV 1040
Db 244 GLAVGTAFNGVDFEGTFHVTVDYDAGFIFGYODSSFYV 287

RESULT 10
P79786 PRELIMINARY; PRT; 181 AA.
AC P79786;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE THROMBOSPONDIN-1 (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97209337; PubMed=90556637;
RA Tucker R.P., Hagios C., Chiquet-Ehrismann R., Lawler J.;
RT "In situ localization of thrombospondin-1 and thrombospondin-3
transcripts in the avian embryo."
RL Dev. Dyn. 208:326-337(1997).
DR EMBL; U76994; AAB19208.1; -
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02412; tsp_3; 5.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 20202 MW; 365E8B0E97BEAE11 CRC64;

Query Match 12.6%; Score 833; DB 13; Length 181;
Best Local Similarity 75.6%; Pred. No. 4.2e-58;
Matches 136; Conservative 22; Mismatches 22; Indels 0; Gaps 0;

QY 686 GDGLICGEDSLDGPWNLNLVCATNATYHCKIDNCPLHNSQGEFDKDGIGDADCD 745
Db 1 GNGIICEEDTDLGWPENLVANATYHCKKDCNCPNLSQGEYDKDGIGDADCD 60
QY 746 NPGVTDEKNCOLLFPNPRQADYKDEVDGDRDNCPCYVHNPAQIDTDNNGEGDACSVDIDG 805
Db 61 DDGIPDRDNCPCFYNPQQYDYDRDDVGDRCDCNCPYHNPNDEITDNNNGEGDACAVIDG 120
QY 806 DDVFNERNRNCPCYVYNVDQRTDGDGVDGDCDNCPLVHNPQDQTVNDLVGDCQDNNEDID 865
Db 121 DGVLERNRNWQYVYNVDQRTDLDGVDGDCDNCPLHNPNQEDTSDSLIGDECDNNQETD 180
```

```
RESULT 11
ID Q15667 PRELIMINARY; PRT; 376 AA.
AC Q15667;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THB51 PROTEIN PRECURSOR (FRAGMENT).
GN THB51.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92278426; PubMed=1350660;
RA Burglin T.R., Barnes T.M.;
RT "Introns in sequence tags [letter; comment].";
RL Nature 357:367-368(1992).
DR EMBL: M14326; AA61237.1;
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; VWF.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF00093; vwc; 1.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00214; vwc; 1.
DR PROSITE: PS01208; VWF; 1.
KW SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 >376 POTENTIAL.
FT NON_TER 376
FT SEQUENCE 376 AA; 41329 MW; 313B62C0D253BA2B CRC64;

Query Match 9.7%; Score 639.5; DB 4; Length 376;
Best Local Similarity 34.4%; Pred. No. 2.7e-42;
Matches 137; Conservative 75; Mismatches 157; Indels 29; Gaps 7;

QY 1 MWRL-VLLALWVFSQHGQDKD-TTFDLFSINIRKKTIGAKQFRGPDGPVPAFRV 58
: ||| : | : : : ||| : : : : ||| : |||
DB 3 LAWGLVFLMHVCTNRIPESGDNVDFIPELTGAARKGSRRLVRGPDPSAPFRIE 62
: ||| : | : : : ||| : : : : ||| : |||
QY 59 RFDYIPPNADLSKITKIMROKEGFFELTAOLKQDKSGRTLLALEGPGLSQRFIYSN 118
: ||| : | : : : ||| : | : | : ||| : |||
DB 63 DANLIPVPDDKFDQLVDVRAERKGLLLASLRQMKRTGTLALERKDHSGQVFSVSN 122
: ||| : | : : : ||| : | : | : ||| : |||
QY 119 GPADTLDTLYIDGTRHVVLESDVGLADSQMKNVTQVAGETSLYHVGCDLIDSPALDEP 178
: ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 123 GRAGTLDSLTVQGHVSVVEALLATQWKSIITLQVEDRAQLYIDCKMENAEIDVP 182
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 179 F---YEHQAESRMVAVKGSARESHFRGLLQNVHLVENSVEDILSKKGQGGGASIN 235
: ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 183 IQSVETRLDLSARLRIAKGGVND-NFQVLQNVRFVGTTPEDILRNKGCSSSTSVLLT 241
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 236 AISENTETLRLGPHVTEVGVSSRRPEVGCERCEELGNMVOELSGHLVLPQSENLK 295
: ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 242 L--DNNVNGSPARTNFIQHTKDLQAICISCELSMVLRLGRLTIVTTLQDSIR 299
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 296 RYSDNQELWELIGPPKTRNMSACWDQGFRAENETWVDSCTCTCKFKFKTICHOITC 355
: ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 300 KYTEENKELANELRRPP-----LCYHNGVQYRNEEWTVDSCTECHQNSVTICKYSC 353
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 356 PPATCASPSFVEGECPCSLHSDVDEEGNSPWAETQC 393
: ||| : | : ||| : | : ||| : | : ||| : | : |||
```

```
Db 354 PIMPCSNATVPDGECCPRC-----WPRC 376

RESULT 12
Q28195 PRELIMINARY; PRT; 128 AA.
ID Q28195
AC Q28195;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE THROMBOSPONDIN-2 (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.";
RL J. Cell. Physiol. 167:164-172(1996).
DR EMBL: X89512; CAA61683.1;
DR InterPro: IPR003129; TSPN.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00210; TSPN; 1.
DR NON_TER 128
FT NON_TER 128
FT SEQUENCE 128 AA; 14400 MW; 9F0B36F9DA5FA82B CRC64;

Query Match 8.8%; Score 580; DB 6; Length 128;
Best Local Similarity 83.6%; Pred. No. 2.8e-38;
Matches 107; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 23 KDTTFDLFSINIRKKTIGAKQFRGPDGPVPAFRVRFYIPPNADLSKITKIMROKE 82
: ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 1 EDTAFDLFSINIRKKTIGAKQFRGPDGPVPAFRVRFYIPPNADLSKITKIMROKE 60
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 83 GFFLTAOLKQDKSGRTLLALEGPGLSQRFIYSNVPADTLDTLYIDGTRHVVLESDV 142
: ||| : | : ||| : | : ||| : | : ||| : | : |||
DB 61 GFFLTASMKQDRSRGRTLLALEGPGATHRQRFIYSNVPADTLDTLYIDGTRHVVLESDV 120
: ||| : | : ||| : | : ||| : | : ||| : | : |||
QY 143 GLADSQWK 150
: ||| : | : ||| : | : ||| : | : |||
DB 121 GLADSQWK 128
: ||| : | : ||| : | : ||| : | : |||

RESULT 13
Q96RW7 PRELIMINARY; PRT; 5636 AA.
ID Q96RW7
AC Q96RW7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMICENTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RT "Human hemicentin gene.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF156100; AAK68690.1;
DR SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;

Query Match 7.6%; Score 505; DB 4; Length 5636;
```

Db 5164 DCDNTIGSYRCVYRCGSGFRRTSDGLSCODINECOESSPCHQRCFNAIGSFHCCEPGYQ 5223

Qy 902 NDG-----VPDDRDN-CRLVFNPDPQBDLGDGRG---DICKDDF---DNDNIPDIIDDVCP 949

Db 5224 LKGRKCMNVNECRQVCR---PDCKKTRGGYKCIDLCPNGMTKAENGTCIDIDE-CK 5278

Qy 950 EN-----NAISDTDFNFQMPLDPKGTTQIDPNWVIHQG-----KEL 988

Db 5279 DGTQCRYNQICENTRSSYRCV--CPRG-----YRSOGVGRPCMDECEQVKKPC 5327

Qy 989 VQTANSPPGIATVGDFDEFGSVDSGTYVTNRDDDYAGFV---FGYOSSREYVVMWKQ 1044

Db 5328 AHQCSNTPG-----SFKCICPPQHLLDGDK--SCAGLERLPNYGTGYSS-YNLARSP 5378

Qy 1045 VTQTYWEDQPTRAYG--YSGVSLKVYNSTT 1072

Db 5379 VRNNYQQQHYROYSHLYSSYS-EYRNSRT 5407

RESULT 14

Q96SC3 PRELIMINARY; PRT; 2673 AA.

AC Q96SC3;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE FBULIN-6 (FRAGMENT).

GN FIBL-6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RS SEQUENCE FROM N.A.

RC TISSUE=MELANOMA;

RA Kostka G., Timpl R.;

FT "Partial sequence of fibulin-6 with a c-terminal region related to domain II and III of the fibulin family.";

RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AJ306906; CAC37630.1; -.

FT NON_TER 1

SQ SEQUENCE 2673 AA; 291017 MW; BEAEC30B8340E272 CRC64;

Query Match 7.6%; Score 504; DB 4; Length 2673;
Best Local Similarity 22.3%; Pred. No. 2.6e-30;
Matches 287; Conservative 147; Mismatches 406; Indels 448; Gaps

Qy 52 VPAYRFVRFDYIYPVNADLSKITKTMRQKEGF-LTAQLKODGKSRGRTLLALEGPGLSQ 110

Db 1338 IPAH----FD---SVANGHSELVIERYSKEDSGTYCTAE-----NSVGFVKAI----- 1378

Qy 111 RQFEIVSNGPADTLDL-TYWID--GPRHVVSLEDVG--LADSOW--KNVTVOVAGETYSL 163

Db 1379 -GFVVYKPEPVFGDYPSSHIEPILGNALINGEVGDDPTPTIQWRKGVDEISHRRQL 1437

Qy 164 HVGCDLDSFALDEPFEHLQAERSMYVAKGSARESHFRGLLQVHLVFNVEDILSK 223

Db 1438 NG-----SLAIYGTNED-AGDYTCVATNEAGVVERSMNLTLSQSPITILEPETVINA 1491

Qy 224 KG-----CQ-----QGGA-----EINAISENTEFLRLGPHVTTEYVGPSSER 261

Db 1492 GKKIILNCQATGEPPQPTITWSRQGHISISWDVRVNLNN-----SLYT--ADAQ 1538

Qy 262 RPVCERSCEELGNWQVELSGLHVLVYNQPSENLKRVSNNDNQFLWLIGLPPKTRMMSACW 321

Db 1539 KEDTSEPEC-----VARNLWG-SVLVRVP-----VIVOVHG----- 1569

Qy 322 QDGRFFAENETWVDSCCTTC-----KPKFTICHQITCP-PATCASP---SFVEGECPCS 373

Db 1570 -----FSQSAW--RACS-VTCGKGLQKSRCLNQ---PLPANGKPKCGSDLEMRNCQN 1618

Qy 374 CLHSVDGEGSWPAEWTCQSCTGSGTGQQRGRSCDVTS-----NTCLGPSIQTRACSLS 428

```
Db 1619 KPCPVDG--SWSEMSIWEBCRSCGRGNOTRTTCNNFNSVQHGRCPEGNAVEIIMCIR 1676
QY 429 KCDTRIDGGHSHWSPSSCSVTGCVGNITRILCNLSPVPMQMGKNCKGSGRETKACOG 488
Db 1677 PCPVH---GAWSAWCPMGCTSESCGKGTQTTRARLGNPPPAFGSGSYCDGAETQMVCNE 1732
QY 489 APCPIDGRWSPWSPSACTVTTCAGGIRTRVCNSPEQYGGKACVGVQVQRMCKKRS 548
Db 1733 RNCPVHGWKATWASALSVSCGGARONTRCSDPVQYGGKCKGSDGVSDFCNSDPC 1792
QY 549 PVDG-----CLSNPCFPGAQCSSFPD----- 569
Db 1793 PTHGNWSPWGWGTCRSCNGSGOMRRYTCNDNPPNSNGRACGPDQIQRNCTDMCPVD 1852
QY 570 ---GSH-----SCGS-----C-PVGFLNGTHCEDLDECALVPDICFSTSKVP 608
Db 1853 GSWGSHWSPWSPSASCSGGGKTRKRLCDHPVPVK-GRPCPG-----DTTQVT 1899
QY 609 RCVNTQPGFHCPLPCPPRYRG-----NCPV 632
Db 1900 RC-NVQA---CPGGQRARGSVIGNINDEFGIAFLNATITDPSNSTRIRAKITNVP 1955
QY 633 GVGLEBAK-----TEKQVCEPENCKDKTHNCKHKAECIYLYGHFSDPMYKCECQ 681
Db 1956 SLGSAMRKIVSLNPIYWTAKKEAVNGFT-----LTNAVEKRETO 1998
QY 682 TGYAGDGLI-----CGESD-----LDGWPNLVLCATNATYHIKNCPLHNSG 727
Db 1999 VEFATGEILQMSHARGLDSDGSLLDIVVSGYV-LQLQSPAETV---VKDY----- 2046
QY 728 QBDPDKDGDGACDDDD---DNDGV-----TDEKDNQ-----LLFNPRQADY 767
Db 2047 TEDIQTGPGQLYATSTRFLTIDGISIPYWNHTVYDQAQGRMPFLVETLHASSVESDY 2106
QY 768 DRDE-----VGRDCNCPYVH-----NPAQIDTDNNGEGDAGSDVIDGDDVFN 810
Db 2107 NQIETLGFKIHASISKGDRSNQCPSGFTLDSVGPFCADDECAAGNPCS----- 2156
QY 811 ERDNCPPYVNTQDRTDGGVGDHDCNCP---LVHNPQT---DVNDLVG-DQCDNNEDI 864
Db 2157 --HSC-----HNAMTYTCSKPKGLTIAADGRTQDIDECALGRTCHAGQDC 2202
QY 865 DD-----DGHNQND---NCPYISNANQD---HDR-DQGD---ACDPDDDD 903
Db 2203 DNTIGSYRCVWRCGSGFRRTSGLSCQDINECEQSSPCQRCFNAIGSFHCCEPGYQK 2262
QY 904 G-----VDDDRN-CRLVFNPDQEDLDGDRG---DICKDF---DNDNIPDIDVCPEN 951
Db 2263 GRKCHDVNECRQVCR---PDHCKNTRGGYKCIDLCPNGMTKAENGTCIDIDE-CKDG 2317
QY 952 -----NAISETDFRNQFVPLDPKGTQIDPNWVIRHQG-----KELVQ 990
Db 2318 THQCRYNQICNTRSYRCV---CPRG-----YRSGVGRPCMDINECEQVFKPAH 2366
QY 991 TANSDFGIAVGDFEGSVDFSTFTVNTDRDDYAGFV-----FGYQSSRFYVVMKQVT 1046
Db 2367 QCSNTPG-----SFKCICPPQSHLLGDGK---SCAGLERLPNYGTQYSS-YNLARFSPVR 2417
QY 1047 QYWEQDPTRAYG---YSGVSLVNVSTT 1072
Db 2418 NNYQPOOHYRQYSHLYSSYS-EYRNSRT 2444
```

RESULT 15

```
Q9TTS5 PRELIMINARY; PRT; 685 AA.
AC Q9TTS5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SCO-SPONDIN (FRAGMENT).
GN SCO-SPONDIN.
```

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-SUBCOMMISSURAL ORGAN;
RA Gobron S., Creveaux I., Monnerie H., Elbitar F.;
RT "Characterization of cattle SCO-spondin.";
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ132106; CAB53759.1;
DR InterPro; IPR002919; TIL.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01826; TIL; 2.
DR Pfam; PF00090; TSP1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS00092; TSP1; 4.
FT NON_TER 1
FT NON_TER 685
SQ SEQUENCE 685 AA; 71389 MW; A003F880E746D5D3 CRC64;
```

Query Match 6.7%; Score 442; DB 6; Length 685;
Best Local Similarity 24.5%; Pred. No. 2.8e-26;
Matches 125; Conservative 39; Mismatches 146; Indels 200; Gaps 19;

```
Qy 320 CWQDGRFAENETWVWDSCTTCKKFKTICHQIT-CPPATCASPFEVEGECPCSLHVS 378
Db 227 CTHGRLHPPGSA-VLRPCENCSC-----VSLITNCTSWPK----- 263
Qy 379 DGEEGSWPAEWTOGCVTCGSGTQGRSC-----DVTSTNCTLGPSTQ 421
Db 264 EGQPTWSPWTPWSECSASCGPARRKHFRCTRPPGAPSSMAPLLLSVPPLCPEAE 323
Qy 422 TRACSLKCDTRIRODGGHSHWSPSSSVTCGVGNITRILCNLSPVPMQMGKNCKGSR 481
Db 324 EEPCLLPECD---RAGMGPMGWSPSSCSRS CGGLRSRACDQPPPGGLGDIYCEGPRA 379
Qy 482 ETQACQ----- 487
Db 380 QGAACQALPCPVNTCTATEGAEYSACGPPCRSCDDLVHCVHWCQPCYCPGQVLSADG 439
Qy 488 -----GA-----PCPIDGRWSPWSP 502
Db 440 TVHVQPGHCSCLDLTLGERHRPGAQLAKPDGCNYCTCSEGLTCTDLPCVPVGAWCPWSE 499
Qy 503 WSACTVTTCAGGIRETRVCNSPEQYGGKACVGVQVQ---RQWC-NKRSQVVDGCLS 555
Db 500 WTACSQPCOGOTRTRSRACSCPAPOHGGAPCPGEGAGEAGAHQRETCASTPECPVDGAW 559
Qy 556 NPCFPGAQCS-----SFPDGSWSGSCSPVGF-----LGNGTHCEDLDECALV 597
Db 560 -PWGSPWCEVCLGRSHRSRECSWPTSEGGRCPPGHRQSRPCOGNSTQCTD---CAGG 615
Qy 598 PDICFSTSKVPR-CVNTQPGFHCPLCPPRYRGNQVGVGLEAAKTEKQVCEPENCKDKT 656
Db 616 QDLLPCGQPCPRSCDLSPGVBC-----QPSMSGCGQQPRCG---CPEGQLSQDGL 662
Qy 657 HNCXKHAECIYLYGHFSDPMYKCECOTGYAG 686
Db 663 --CVTPSQC-----RCQYQPGAMG 679
```

Search completed: August 9, 2002, 10:06:10
Job time: 291 sec